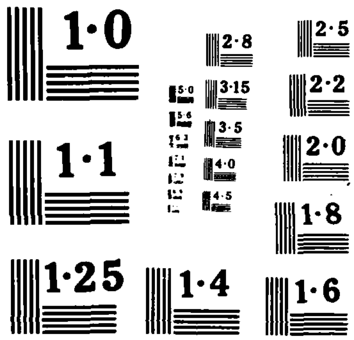
AD-A187 297 GRAPHIC EXPLORATORY ANALYSIS OF VARIANCE ILLUSTRATED ON A SPLITTING OF THE JOHNSON AND TSAO DATA(U) PRINCETON UNIV NJ E G JOHNSON ET AL. 1987 ARO-23360.7-MA 1/1

UNCLASSIFIED DAAL03-86-K-0073 F/G 12/3 NL

END
DATE
FILMED
2 88


1·0
2·8
2·5
5·0
3·15
2·2
5·6
3·5
2·0
4·0
1·1
4·5
1·8
1·25
1·4
1·6

UNCLASSIFIED MASTER COPY - FOR REPRODUCTION PURPOSES

FILE COPY

2

SECURITY CLASSIFICATION OF THIS PAGE

AD-A187 297

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARO 23360.7-MA
6a. NAME OF PERFORMING ORGANIZATION Princeton University; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code) Princeton University, Princeton, NJ 08540	7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211, Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAL03-86-K-0073
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211, Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Graphical Exploratory Analysis of Variance Illustrated on a Splitting of the Johnson and Tsao Data

12. PERSONAL AUTHOR(S)
Eugene G. Johnson, John W. Tukey

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Reprint	FROM ____ TO ____		

16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

DTIC
ELECTE
OCT 2 8 1987
S D
D

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ARO 23360.7-MA

Copyright © 1987 by John Wiley & Sons, Inc.

CHAPTER 10

Graphical Exploratory Analysis of Variance Illustrated on a Splitting of the Johnson and Tsao Data[1]

Eugene G. Johnson
Educational Testing Service

John W. Tukey
Princeton University

DTIC COPY INSPECTED 4

A-1

INTRODUCTION

Since our main purpose is to illustrate how a combination of graphic display and simple arithmetic can be used to enhance the effectiveness of Daniel's half-normal plots, we shall focus on an analysis of a ^ × 4 × 7 data set provided by the responses of person IB1 in the "methodological experiment" presented by (Palmer) Johnson and Tsao (1944) and also analyzed by Palmer Johnson (1949) and by Green and Tukey (1960). The full data set involves 8 persons. We will return below to a fuller description of the data set and a discussion of how it might be analyzed, and elsewhere to a discussion of how the analysis might be carried further, with special attention to the identification and sterilization of exotic values.

We need to make clear our overall prejudices and purposes — a (well-founded) prejudice that most analyses of variance with 3 or more factors are used for exploration rather than confirmation, with a

1 Prepared in connection with research at Princeton University sponsored by the U. S. Army Research Office (Durham).

clear description of the apparent behavior being much more important than formal significance tests, and a clear purpose to reach as simple and complete description of the data behavior, in the instance before us, as we know how, taking reasonable account of questions of multiplicity, but not overemphasizing precise significance. We want a description of the *appearance* of our data, even though more data might not confirm that appearance.

Since we further believe that good techniques come from the accretion of many ideas, not just from a single brain wave, we are not dismayed by the appearance of at least 12 conceptual ingredients (3 old, 1 due to Daniel). Rather we wonder where the 13th and 14th will come from. To tease the reader's imagination, we list the 12 ingredients so far at hand (the later ones need not be as large or important as the earlier):

1) classical analysis of variance
2) aggregation
3) half-normal plotting
4) horizontalized plotting
5) scission into bouquets of contrasts
6) pretrimming by nomination
7) post-trimming by election
8) nominated bouquets
9) 2nd order trimming (super-election)
10) reformulating a response
11) rethinking a scission
12) refactoring an analysis

We believe that these ingredients can be used in any factorial analysis of variance — and in many others of different form.

The basic elements underlying all this are:

(A) basic ANOVA concepts of decomposition — of separating of each number into parts, each part coming as purely as we can arrange from its specified "source,"

(B) anticipation of revision in the light of the data, not only of numbers but also of the style and form of separation,

(C) use of long-term insight to select specifics for trial,

(D) use of pictures to see what may need special treatment or modification,

(E) use of arithmetic to conduct modifications,

(F) ultimately a combination of (a) numerical summaries, hopefully depictable, and (b) pictorially apparent absence of what else might plausibly be present.

We now discuss the ingredients, and explain their concatenation and mixing, in terms of the single 2 × 4 × 7 data set mentioned above, showing how they lead to a reasonably compact description of the 56 numbers.

CONTENTS

PART A: ANALYZING IB1's PERFORMANCE IN GRAMS

A1. Looking at the Data Overall

The 2 × 4 × 7 data set for person IB1 (one of the two blind males in the full experiment) involves 2 dates (1,2), four rates (50, 100, 150 and 200 grams per 30 seconds), and seven (initial) weights (100, 150, 200, 250, 300, 350 and 400 grams). The experimental procedure involved attaching a pail by a lever system to a ring on the subject's finger. One of the seven initial weights was placed into the pail, and then water was allowed to flow into the pail at one of four constant rates until the subject reported a change in pull on the finger. The intended response, the difference limen (D.L.), was measured by the amount of water added by time of report. Five determinations were made for each of the 28 rate-weight combinations, and the average of these values was used as the response. The entire experiment was conducted, for each person, at each of two dates, one week apart.

The full experiment consisted of 8 persons, two persons in each cell of a 2 × 2 design for male vs. female and sighted vs. blind. In their analysis of the complete data set, Green and Tukey noted that person IB1 had a pattern of response that was considerably different from that of the other persons (including the other blind male) and designated him as the "eccentric blind man." For this very reason, we have selected person IB1 for our initial (within person) analysis.

TABLE 1. Average difference limen in grams for person IB1 – male, blind.

Rate		Initial Weight (Grams)						
(gm/30 sec)	Date	100	150	200	250	300	350	400
50	1	24.2	25.3	25.1	17.6	20.7	19.4	17.3
	2	41.2	29.8	28.5	23.8	20.9	17.8	13.4
100	1	48.1	41.2	31.4	30.4	39.9	36.7	35.5
	2	59.1	59.7	48.7	38.1	30.7	28.4	27.2
150	1	60.9	52.0	58.2	60.6	57.1	57.9	49.5
	2	75.8	79.9	69.1	64.4	42.2	53.1	36.3
200	1	69.9	76.7	82.4	76.4	71.4	76.9	79.6
	2	148.3	123.1	73.5	61.9	77.8	56.0	53.2

As a first step in the analysis of the performance of IB1 to the experiment, we display his responses. The actual values appear in Table 1; a graphical display of the responses is shown in Figure 1. Figure 1 consists of a series of 7 plots, one for each level of weight (indicated on the horizontal axis). Each of the seven plots shows the relationship between the response (D.L. in grams) and rate for both dates. The responses for each date are connected by a broken line. In considering Figure 1, the first thing that strikes the eye is the strong linear relationship between D.L. and rate. In fact, such a relationship, which was also found by Johnson and Tsao and by Green and Tukey, holds for each person. In Part C, we will consider an analysis of the data for person IB1 which uses a different dependent variable (log(response time)) and which produces a particularly simple interpretation of the relationships between level of response and the various factors. For the moment, however, we will press forward with an analysis of person IB1 with response in the original units — D.L. in grams.

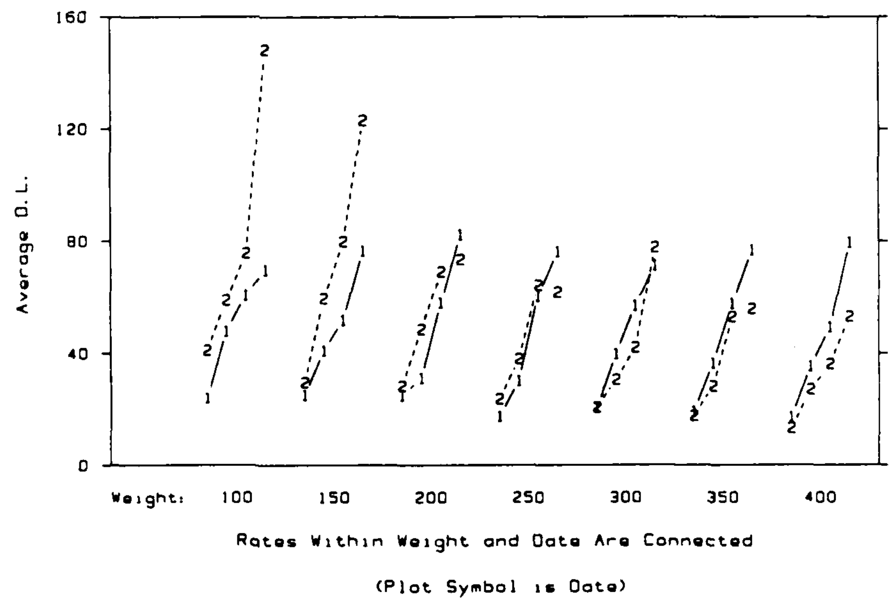


Figure 1. Person IB1: male, blind. Average difference limen (grams).

Crude Classical Analysis of Variance

A standard analysis of variance table of the $2 \times 4 \times 7$ factorial design for person IB1 with difference limens in grams as the dependent variable is given in Table 2. In the table, we have denoted the independent variables as R for rate, W for weight and D for date. In determining the values for "F" and the significance levels, the line of the table corresponding to the three-factor interaction DRW was (naively) taken as measuring appropriate error for each of the other lines (extreme model 1).

Given the strong linear relationship between D.L. and rate within each date-by-weight combination noted from Figure 1, it is not at all surprising to find that the largest mean square is that associated with rate. The only other "significant" lines in Table 2 are the weight main effect and (somewhat less strongly) the date × weight two-factor interaction.

Crude Aggregated Analysis of Variance

Using the sort of aggregation proposed by Green and Tukey (and described in section A9), the result would be as in Table 3.

We shall consider more refined analyses starting from (analogs) of these two tables. The results are similar to those for the classical analysis, although the 2×7 date and weight table is reassembled (from D, W and DW).

Single df's

A (traditional) [aggregated] analysis would now proceed to pull apart from the various (significant) [remaining] lines one or more single-degree-of-freedom components, comparing the magnitudes of each of the constituent contrasts with that of (the three-factor interaction mean square) [an appropriate aggregated denominator] taken to represent error. Rather than beginning with such an approach, we will cut up each of the lines of the (full) [aggregated] analysis of variance table (including the DRW line) into single degree-of-freedom components — contrasts — and compare the sizes (absolute values) of these contrast components graphically by a technique related to Daniel's half-normal plots.

In such an analysis, to be described shortly, we do not assume that any prechosen line of any analysis of variance table is necessarily solely measuring error. Rather, we assume that any of the selected contrasts which combine to constitute some line (i.e., a main effect, a two-factor interaction, ..., a n-factor interaction) might have mean values different from zero, but that most of the totality of them will serve for error estimates. That is, we assume the bulk of the (properly defined) single-degree-of-freedom components are actually measuring error — or come close to doing so — but we do not *a priori* specify which ones, or which error.

We selected our 2 × 4 × 7 subset of the data in such a way as to avoid major complications with multiple error terms, which arise when the whole data set is considered. Similar techniques will apply to factorial data sets deserving more error terms.

TABLE 2. Standard analysis-of-variance table for person IB1 (dependent variable is difference limen in grams).

Source	df	MS	DEN	F	Sig
D	1	348	149	2.33	not
R	3	8514	149	57.06	0.01%
W	6	772	149	5.18	0.5%
DR	3	21	149	0.14	not
DW	6	545	149	3.65	2.5%
RW	18	74	149	0.50	not
DRW	18	149	–	–	–

TABLE 3. Aggregated analysis-of-variance table for person IB1.

Label	df	MS	DEN	F	Sig*
Rate	3	8514	105**	81.10**	0.01%
Date and Weight	13	635	105	6.05	0.01%
Residual	39	105	–	–	–

* Notice that (a) significance levels of F-values require a large, rather unspecified multiplier for multiplicity and (b) 0.01% is the most extreme level considered.

** Would have been 635 and 13.4, respectively, had any rate interaction appeared in the following (date and weight) aggregation.

A2. Bouquets of Contrasts

Since we are going to focus on single degrees of freedom, we need to break up each factor into a bouquet of contrasts, each a single degree of freedom. We expect to use the natural combinations (outer products) of these one-factor contrasts to also break up the two-factor and three-factor interactions into single-degree-of-freedom contrasts.

Because the two factors to be broken up (date is already a single contrast) are scales with equispaced versions (levels), it is perhaps natural to consider the classical orthogonal polynomials as a possible first choice for the basic bouquet. Besides these contrasts there are other types of orthogonal contrasts which, depending on circumstances, may have greater utility. We shall return to some of these in section D2.

If the versions of our factors had been only ordered, not measured, we might have followed Abelson and Tukey (1963) in selecting an initial contrast, or conceivably, have separated the response into monotone increasing and monotone decreasing parts. Unordered versions can often be sensibly partitioned, although we may have to be somewhat arbitrary in some or all of our choices of contrasts.

Various bouquets of contrasts have their place in the analysis of data. We will use a number of different bouquets in our analysis of our $2 \times 4 \times 7$ data set. Given a collection of bouquets of single-degree-of-freedom contrasts, one for each line of the standard analysis of variance table, the next step in the continued analysis of the data is the assessment of what the values of the contrasts are trying to tell us about the relationship within and between the various factors. This assessment will be done graphically via a procedure related to Daniel's half-normal plots.

A3. Horizontalizing Plots

The classic "half-normal plot" relates the sizes (absolute values) of the normalized contrasts, ordered by size, with typical values of order-statistics of the half-Gaussian distribution by plotting (ordered) size of contrast versus typical order-statistic. The corresponding natural reference is a line *through the origin,* whose *slope* corresponds to an estimate of the underlying scale σ. To make internal comparison much easier, we shall instead plot

$$\frac{\text{size of contrast}}{\text{typical order statistic}} \quad \text{versus} \quad \text{(typical order-statistic)}$$

thus making the natural reference a *horizontal line*, whose *height* corresponds to an estimate of σ.

Display Ratios

While classical probability plots often use the unit deviates corresponding to $\frac{i}{d+1}$ or $\frac{i-½}{d}$ as the typical order statistics for a sample of d values, we shall work very close to the order-statistic medians by using deviates corresponding to $\left(i-\frac{1}{3}\right) / \left(d+\frac{1}{3}\right) = (3i-1)/(3d+1)$. For the half-Gaussian (the distribution of the positive square root of any chi-square on one degree of freedom), this means using the half-Gaussian working values, the ith such (of d) being

$$c(i:d) = \Phi^{-1}\left[½+½\left(\frac{3i-1}{3d+1}\right)\right] = \Phi^{-1}\left(\frac{3i+3d}{6d+2}\right)$$

where Φ is the unit Gaussian cumulative distribution function.

Thus, we shall plot the

$$\text{display ratio} = \frac{|C(i:d)|}{c(i:d)}$$

versus the typical order statistic $= c(i:d)$ where $|C(i:d)|$ is the ith largest size of our contrasts. Under the simple (null) model that the sizes of contrast $|C(1:d)|, \ldots, |C(d:d)|$ represent a set of order statistics from a sample of size d from a half-Gaussian distribution with scale σ and location 0, each of the d display ratios provides an estimate of σ.

We compute the display ratios separately for each bouquet of contrasts, one bouquet for each line of a conventional analysis of variance table. A plot of display ratio versus working value for a particular bouquet shows:

1) the general level of variability, hopefully background noise, captured by a typical defining contrast of the bouquet and measured by a horizontal line, and

2) the relative magnitude of the various sizes of contrast in terms of the general level for that bouquet and in terms of what would be expected under a simple (null) model.

We tend to focus on the display ratios for the largest sizes of contrasts, interpreting relatively large display ratios as indicating potentially meaningful contributions, likely to be worth separate description.

A plot of display ratio versus working value for a bouquet will sometimes produce slightly confusing appearances, when granularity, arithmetic errors, or other causes of individual exotic values keep the contrasts of smallest size from being as small as a simple model suggests they ought to be. Thus, relatively high values of the display ratio for quite small working values should often be ignored. If considered, they should usually be regarded as suggesting isolated errors, exoticities or granularities. (We turn later (elsewhere) to looking for such isolated phenomena.) A general downward trend (to the right) invites similar interpretation and treatment.

Although we compute the display ratios separately for each bouquet, ordinarily we will overlay the plots for each bouquet of a given type (main effects, 2-factor interactions, etc.) on the same figure, connecting the points for contrasts in each bouquet by a broken line. This allows both for internal comparison of the sizes of contrast within a bouquet and comparison with the sizes of contrasts of the other bouquets of that type. By using the same vertical scale for all of the plots, we can also compare the sizes of contrasts across the various types. The latter comparison allows the assessment of the relative importance of a given contrast in the experiment as a whole and also, by comparing the general level of one bouquet with that of the others, indicates if the set of defining contrasts for a given factor might be replaced by another set of defining contrasts to produce a simpler account of the data. Such a possibility exists if the general level of a particular bouquet, particularly a main effect bouquet, is above the levels of the other bouquets. We will discuss the possible causes of this in Part D.

A4. Display Ratios in the Example

We now return to subject IB1 and apply the above procedure, using the polynomial contrasts. The result is shown in Figure 2, where we have grouped the bouquets into three sets, plotting the three main-effect bouquets together in the first panel, the three two-factor-interaction bouquets together in the second, and the three-factor interactions in the third panel. The vertical scale for the three plots is the same, allowing for the comparison of magnitude of the display ratios for all bouquets.

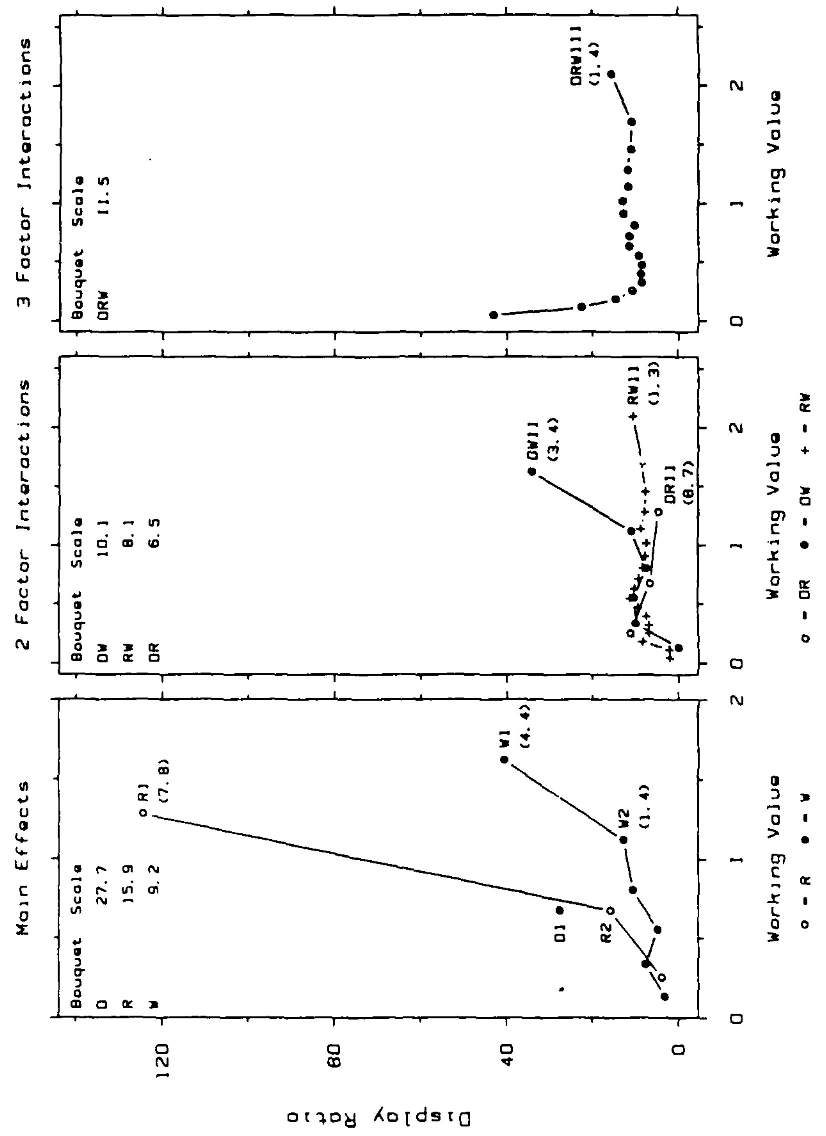


Figure 2. Person IB1 — D.L.in grams. Polynomial contrasts.

The points for contrasts within each bouquet are connected by a broken line, and the largest size of contrast within each bouquet is labeled. Occasionally, other contrasts of large size will be labeled. The notation for the various bouquets is as in the initial analysis of variance (Table 2); the order of polynomial contrast is indicated by a number following the bouquet label. For example, *R* 1 is the linear contrast for rate, *W* 2 is the quadratic contrast for weight, *DW* 12 is the linear (in date)-by-quadratic (in weight) two-factor contrast, and *DRW* 123 is the linear (date)-by-quadratic (rate)-by-cubic (weight) three-factor contrast. (We shall return to the various numbers attached to the bouquets and to the high individual points in section B1.)

The first thing to note from the plots is that the bulk of the contrasts have display ratios at a level of about 10 grams. In light of this background level, we have to recognize the contrasts with the largest display ratios — *R* 1 (linear in rate), *W* 1 (linear in weight), *DW* 11 (linear in date by linear in weight), *D* 1 (linear in date), and *DRW* 111 (linear by linear by linear three-factor interaction), in decreasing order — as worth careful consideration. The values of these display ratios, along with those for certain of the next largest contrasts within each bouquet, are given in Table 4.

In view of the proportional relationship between D.L. and rate suggested by Figure 1, it is not surprising that *R* 1 is the strongest observed relationship. It is striking that the five largest contrasts in

TABLE 4. Values of display ratio for the largest contrasts within each bouquet for the polynomial decomposition of IB1 (D.L.).

Contrast	Display Ratio
R 1	124
W 1	41
DW 11	34
D 1	28
DRW 111	16
R 2	16
W 2	13
all others	< 11

terms of display ratio are solely composed of linear (straight-line) contrasts and their products. In fact the linear (or linear by linear, etc.) comparisons have the largest size of contrast within every bouquet. We will return to this in the next section.

For the three-factor-interaction contrasts, the display ratios for the contrasts of smallest size are relatively larger than might be expected. As mentioned previously, this phenomenon might be indicating granularity due to rounding or small arithmetic mistakes (one deviant observation tends to contribute similar amounts to each single degree-of-freedom contrast).

Another phenomenon to notice is associated with the *R*, *W* and *DW* bouquets. In each of those bouquets, the largest size of contrast is relatively much larger than the remaining sizes of contrast within the bouquet. Notice, in each of those three bouquets, that the next-to-largest size of contrast is also somewhat high in terms of display ratio, the point for the contrast appearing above the display ratios of all the remaining contrasts of the bouquet. We call the underlying phenomenon "dragging upward" and will discuss it in section A8.

A5. The Largest of the Three-Factor Contrasts

The right-most point for *DRW* requires careful discussion. It seems to continue, and enhance, a general upward trend for the ratios plotted for the contrasts in this bouquet. However, it does not rise far above the others. Had this, for instance, been a linear-by-quadratic-by-quartic three-factor interaction, or some other nondescript contrast among the $1 \times 3 \times 6 = 18$ in this bouquet, we would not have been likely to attend to it. It is, however, the linear-by-linear-by-linear contrast, *a priori* the most distinctive and most likely (however unlikely) to contain something meaningful. To have it come out as the highest absolute value of all 3-factor contrasts is thus, by itself, significant at $1/18 = 5.56\%$, so that it needs at most a little extra push to be worth our honest attention.

Granting, then, that it may include a real effect, how should we interpret it? It is, after all, a three-factor contrast in a situation where the constituent single-factor contrasts are all large. For the present, then, we may not be too wrong to think of it as "probably real, but likely to be a spill-over from the large main effects because of something resembling not-quite satisfactory expression of the response."

A6. Pretrimmed Bouquets; Nomination

Having found linear, linear-by-linear and linear-by-linear-by-linear contrasts outstanding in our example, we must ask ourselves: "In such a situation, where a few contrasts are distinguished above all others in their respective bouquets, why did we not plan to treat them separately in the beginning — not only in this example but in general?" No good answer is available. So let us trim our bouquets — and even pretend that we pretrimmed them in this example — moving the linear contrast out of each single-factor bouquet, the linear-by-linear ("linear-to-the-2") contrast out of each two-factor bouquet, and the linear-by-linear-by-linear ("linear-to-the-3") contrast out of the three-factor bouquet.

Each original bouquet, corresponding to a line in the analysis of variance table and consisting of d contrasts, has now been partitioned into two bouquets:

- a nominated contrast consisting of the single linear-to-the-j contrast; nominated *a priori* as likely to be interesting
- a trimmed bouquet consisting of the $d-1$ remaining contrasts, which collectively are telling us about the contribution of the corresponding line of the analysis of variance table after eliminating variation describable by the linear-to-the-j contrast.

In our example this creates 13 bouquets from the original 7 (since D is already linear-to-the-1, there is no trimmed bouquet for D). (For a related use of the word "nominated" see S. C. Pearce, 1953 or 1976.)

By nominating $D1$, $R1$, $W1$, $DR11$, $DW11$, $RW11$, and $DRW111$ as *a priori* interesting contrasts we have agreed to treat each of these contrasts not as one of the d members of the original effect bouquet, but rather as a separate thing unto itself. As such, we display them using the working value $c(1{:}1) = .674$ to compute the display ratios for each of the nominated contrasts.

By pretrimming our bouquets, removing the nominated contrast from the initial bouquet with d members, and producing a trimmed bouquet with $d-1$ members, we have agreed to treat the contrasts in the trimmed bouquet as collectively separate in impact from the nominated contrast. As such, we assess the magnitudes of the sizes of contrast for the $d-1$ members of the trimmed bouquet in terms of what would be expected from a sample of size $d-1$ from a half-Gaussian distribution and so use the working values $c(1{:}d-1), \ldots, c(d-1{:}d-1)$ to compute the display ratios.

Results in the Example; Nominated Contrasts

Proceeding in this manner with our example data, we produce Figure 3. In order to better show the detail for the trimmed bouquets, we have truncated the vertical axis of the plots at 50 and thus do not show directly the display ratios for the three largest nominated contrasts: *R* 1, *W* 1 and *DW* 11, each of which should be plotted at working value $c(1{:}1) = .674$.

On examining the plots, we see first of all that the display ratios of all the nominated contrasts, with the exception of *DR* 11, are notably larger than the display ratios of any of the contrasts remaining in the trimmed bouquets (with the exception of the display ratio of the smallest size of contrasts in the trimmed three-factor bouquet, corresponding to the (small) contrast *DRW* 133 — possible reasons for this large display ratio have been previously mentioned).

The display ratios for the 7 nominated contrasts are shown in Table 5, both for the nominated contrasts plotted in Figure 3 and, for comparison, as parts of the original effect bouquets plotted in Figure 2.

We can see from the table (and from the plots) how much the display ratios for 6 of the nominated contrasts have each increased when treated as single-contrast bouquets over the display ratios for the same contrasts when treated as the largest member of one of the original bouquets. (The median display ratio, a natural background level, has fallen from 10 to 9.) The ratio of nominated display ratio to original display ratio appears as the last column of Table 5. We also see that, in terms of size of display ratio, the ordering of the display ratios for the nominated contrasts is essentially the same as before, with the exception of *D* 1, which has moved down from the 4th largest to the 6th largest (after nomination).

These increases in the values of the display ratios reflect the sizes of the working value used to compute the display ratios, which have decreased from $c(d{:}d)$ to $c(1{:}1)$. The ratio of $c(d{:}d)$ to $c(1{:}1)$ is exactly equal to the proportional increase in display ratio due to pretrimming. We can see from the table that the ratio of increase grows somewhat as the size d of the original bouquet grows. This growth explains the increase in relative importance of *DRW* 111 and *RW* 11, both of which belonged to bouquets with 18 members. This also helps to explain why *DR* 11, whose display ratio is inflated by a factor of only 1.9, remains at the level of background variability. The plot of display ratio vs. working value for the full *DR* bouquet in

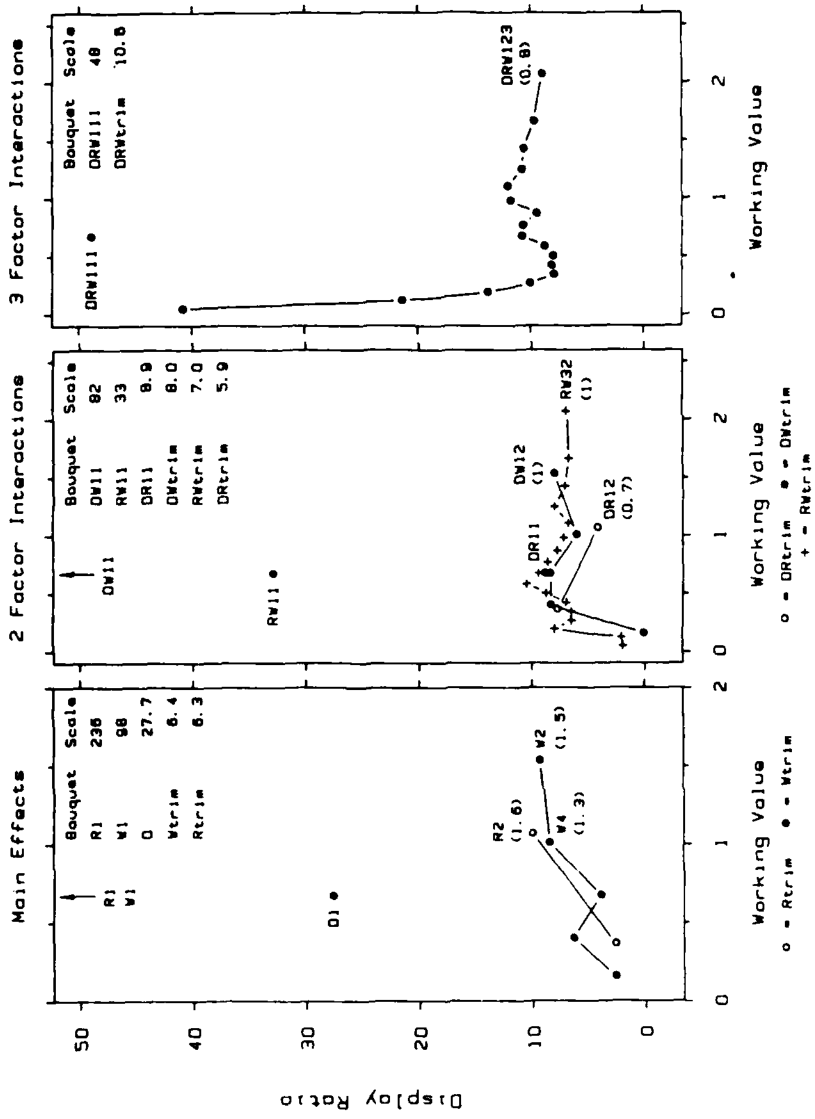


Figure 3. Person IB1 — D.L. in grams. Polynomial contrasts — pretrimmed bouquets.

TABLE 5. Display ratios and reference working values for the nominated contrasts when trimmed out and when left in the original effect bouquets.

	Nominated Contrast		(In Figure 2)			
Contrasts	Display Ratio	Working Value $c(1{:}1)$	Bouquet Size d	Display Ratio	Working Value $c(d{:}d)$	Ratio Of Display Ratios
R 1	237	(.674)	3	124	(1.282)	1.9
W 1	98	(.674)	6	41	(1.620)	2.4
DW 11	82	(.674)	6	34	(1.620)	2.4
DRW 111	49	(.674)	18	16	(2.093)	3.1
RW 11	33	(.674)	18	11	(2.093)	3.1
D 1	28	(.674)	1	28	(.674)	1.0
DR 11	9	(.674)	3	5	(1.282)	1.9
Median Display Ratio Over All 55 Contrasts	9			10		.9

Figure 2 shows *DR* 11 at the end of a general decline; *DR* 11 is relatively smaller than might be expected for the largest order statistic of a sample of size 3.

Before examining the trimmed bouquets in Figure 3, we should reiterate that we are, in this section, discussing only pretrimming. By coincidence (perhaps), the nominated contrasts in our example all were the largest representatives of their respective bouquets. Since the linear-to-the-*j* contrasts are the most easily interpretable (and in general the strongest in many experiments), we should have nominated them *a priori*, in any event. The increase of display ratio on nomination will, of course, be less when the nominated contrast is not the largest in its bouquet.

Other Alternatives

If we had used a different set of orthogonal contrasts other than the orthogonal polynomials to define a bouquet, we could, and probably should, still pretrim whenever one of the contrasts is naturally *a priori* distinguished above all others. It is, or course, also possible to post-trim a bouquet, electing for removal the largest (or largest few) contrasts which attract our attention because of their relatively large display ratios. We discuss this possibility in section B3.

A Nominated Bouquet?

We have nominated 7 contrasts, one from each line of the basic analysis. So far we have treated them as 7 one-contrast bouquets. But why should we not treat them as 1 seven-contrast bouquet, the nominated bouquet? Table 6 shows the display ratios for the nominated bouquet; the first panel of Figure 4 is the corresponding horizontalized plot.

Both Table 6 and Figure 4 show $R1$ as standing out. It might be reasonable to super-elect (see section B3) $R1$ and post-trim the nominated bouquet, separating $R1$ into its own 1-contrast bouquet and leaving the remaining contrasts in a 6-contrast bouquet. The result of this is shown in the last two columns of Table 6 and the second panel of Figure 4. We observe that, whether we post-trim or not, the display ratios of the other 6 nominated contrasts are surprisingly similar. We will return to this point below.

A7. Trimmed Bouquets in the Example

We now return to the trimmed bouquets to consider the effect of nomination and trimming on their display ratios. A consequence of pretrimming can be seen by comparing the plots of the display ratio versus working value for the trimmed and original bouquets. Considering, for example, the three-factor bouquet, we can see from

TABLE 6. Display ratios for the nominated contrasts as members of the seven-contrast nominated bouquet.

	7-Contrast Bouquet		After Super-Electing $R1$ and Post-Trimming	
Contrast	Display Ratio	Working Value	Display Ratio	Working Value
$R1$	94	(1.691)	237	(.674)
$W1$	55	(1.208)	41	(1.620)
$DW11$	61	(.908)	49	(1.119)
$DRW111$	49	(.674)	41	(.804)
$RW11$	47	(.472)	40	(.555)
$D1$	64	(.288)	56	(.336)
$DR11$	52	(.114)	45	(.132)

Figure 2 that the display ratio of the second largest contrast, *DRW* 123, appears at essentially the same level as that of the third largest contrast *DRW* 112. Turning to Figure 3, we now see that the display ratio of *DRW* 123 (now the largest size of contrast) is noticeably lower (by .7 units) than that of *DRW* 112 (now the second largest size of contrast). Looking further, comparing the plots of the trimmed bouquets with the original bouquets, we can see a general tendency for the slopes of the lines between adjacent points to become more negative.

Both this and the reduction in the general level of the display ratios for the trimmed bouquets are consequences of a reduction in the "dragging upward" phenomena to be discussed in the next section.

The Changing Typical Size of Residuals

Starting to act as if we had nominated all linear-to-the-*j* contrasts will change the typical sizes of the display ratios, decreasing such sizes when the nominated contrasts are large, as in this example, and

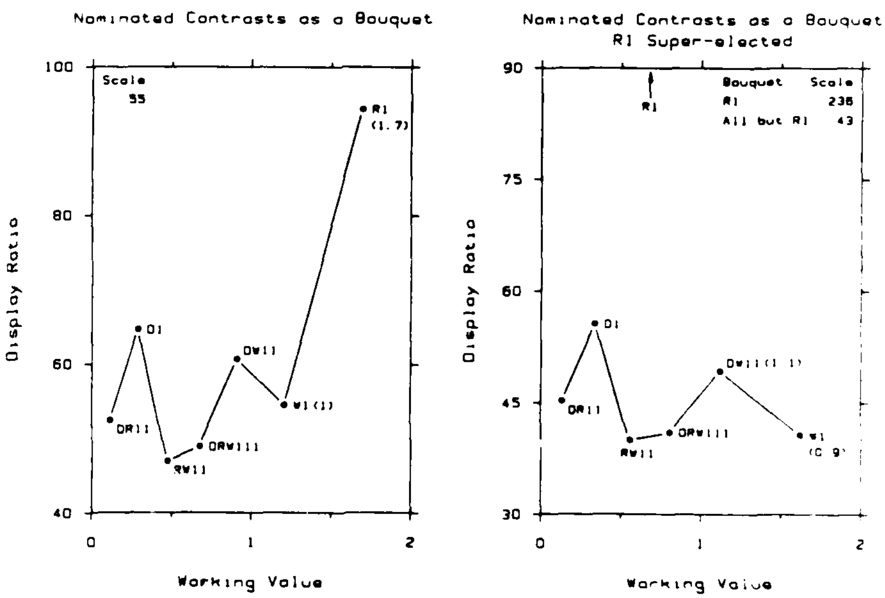


Figure 4. Person IB1. D.L. in grams.

fluctuating them irregularly when the nominated contrasts appear similar in size to the others in their bouquet. The most interesting summary sizes of the display ratios seem to be:

--- before nomination ---	
median for all 55 initial contrasts =	9.9
median for 45 initial non-main-effect contrasts =	9.3
--- after nomination ---	
median for all 55 contrasts with nomination =	8.5
median for all 48 unnominated contrasts =	8.1
median for all 41 unnominated, non-main-effect contrasts =	8.1

Roughly speaking, nomination reduced the median size of display ratio by 13%. (The effect would have been larger if we had not used the median, a highly resistant summary. It is here neither important nor wholly negligible.)

A8. Dragging Upward

In our example (section A4) we saw that retaining the largest contrast in each bouquet tended to make the second-largest contrasts look more distinctive than need be. Do we expect this in general?

We need only look at the divisors, at the order-statistic typical values, to see how this occurs. For $d = 7$ (say before setting the largest apart) and $d = 6$ (after) we have the values in Table 7. On a relative basis, the divisor for what was initially the second largest size of contrast was about ¾ as large before setting aside the largest one as it was after this. It is, in fact, always the case that the typical value of the ith order statistic of a sample of size d, $c(i:d)$, is smaller than $c(i:d-1)$, the typical value of the ith order statistic of a sample of one less. The relative difference is most pronounced for smaller bouquet sizes and is largest for the next-to-largest contrast of a bouquet. Table 8 shows the reference values for the next-to-largest contrast within an original bouquet both before and after trimming out the largest contrast, for the bouquet sizes of our example. The table also includes their ratios $c(i:d)/c(i:d-1)$, and the median value of these ratios.

If we have pretrimmed our bouquets, then the various display ratios for the trimmed bouquets will be reduced from what they would otherwise have been in the original bouquets by fractions indicated by Tables 7 and 8. If the nomination was done in advance of seeing the data — and, also, to a practical approximation, if it was done before any detailed analysis of the data was made — then the after-trimming display ratios will almost surely be more appropriate. (True post-trimming requires somewhat more careful thought.) The larger display ratios before trimming were "dragged upward" by being taken as less exalted order statistics than they deserved to be — because the even higher contrast, deserving of nomination, unfairly seized the highest position. By trimming, we have prevented this dragging upward, restoring the display ratios to what they ought to be.

TABLE 7. Illustration, for $d = 7$, of dragging upward via denominators.

i	$d = 7$ $c(i{:}7)$	$d = 6$ $c(i{:}6)$	Ratio	Ratio to Median
7	1.691	—	—	—
6	1.208	1.620	.745	.882
5	.908	1.119	.812	.961
4	.674	.804	.838	.992
3	.472	.555	.852	1.008
2	.288	.336	.859	1.017
1	.114	.132	.863	1.021
			median = .845	

TABLE 8. Dragging upward via denominators of the next-to-largest contrast.

d	$c(d-1{:}d)$	$c(d-1{:}d-1)$	Ratio	Median Ratio*
3	.674	1.058	.631	.661
6	1.119	1.534	.729	.823
18	1.691	2.070	.817	.937

* Median is of $c(i{:}d)/c(i{:}d-1)$ for $i = 1, ..., d-1$.

A Possible Initial Plot

We could try to reflect this dragging upward effect in an initial plot, at the cost of making things rather "busy" looking. Figure 5 shows the three-factor sizes of contrast with

$$\frac{|C(i:d)|}{c(i:d)} \quad \text{plotted as 0}$$

$$\frac{|C(i:d)|}{c(i:d-1)} \quad \text{plotted as 1}$$

and

$$\frac{|C(i:d)|}{c(i:d-2)} \quad \text{plotted as 2}$$

with the three for each i connected by broken lines. The vertical axis has been truncated at 20 to show the main detail. We see a consistent

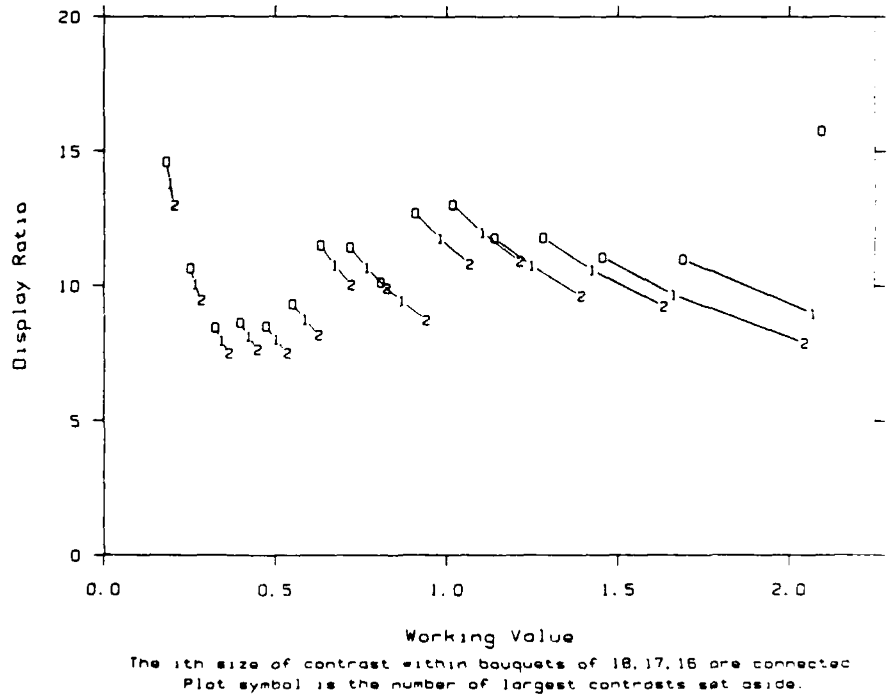


Figure 5. Effect of undragging for the three-factor contrasts. IB1 — D.L. in grams — polynomial contrasts.

decrease in the value of the display ratio as we move from a bouquet of all 18 to a bouquet of the smallest 17 to a bouquet of the smallest 16 sizes of contrast. The horizontal spread in the plot of each size of contrast, indicated by the pair of lines connecting a 0, 1 and 2, shows the relative change in value of the reference value as the bouquet size is reduced. The change in working value versus bouquet size is the most pronounced for the largest size of contrast, as is the decline in the value of the display ratio.

Since we must expect some such effect, even if the largest contrast did not deserve to be nominated, it is not easy to argue cogently from such a plot. We shall not pursue its possible use further.

A9. The Effect of Aggregation

Sections A2 to A8 have been concerned with either (a) choosing a general approach or (b) illustrating the consequence of using the approach starting from Table 2 or an analog. It is time to ask what changes we expect when we start from an analog of Table 3, if we aggregate before going to the individual contrasts.

Opening Analysis

If we are to pretrim, we should choose the nominees before any detailed analysis of the data. So it is natural for us to begin with a conventional post-nomination analysis-of-variance table, involving 13 lines. Table 9 sets out the numbers. We will use the notation "$X(n)$" for the nominated portion of the bouquet labeled by "X".

A striking thing to note from the post-nomination analysis-of-variance table is that the value of the mean square for *DRW*trim, the three-factor interaction after removing the linear-to-the-3 component, is larger than the values of the mean squares for any of the other "trimmed" lines in the table.

The Notion of "Above"

In a general context, one line of an analysis-of-variance table is "above" another line if variability in the "lower" line inevitably penetrates into the "upper" line, a situation often formalized (usually satisfactorily) as: "the expected mean square of the former ("upper") line contains all the terms in the expected mean square of the other ("lower") line."

In doing aggregation after nomination, we need to be somewhat careful in defining what is above what. We will take the view here that "$X(n)$" is above "Xtrim" and also above anything that "Xtrim" is above (using the conventional definition to decide the partial orderings for the "trim" lines). Whether any nominated contrast can be above any other nominated contrast is a consequence of the particular sets of orthogonal contrasts used for the various bouquets. In our particular situation, where we are using sets of orthogonal polynomials (and their outer products) to define the various bouquets, Interpretation One holds that no nominated contrast can be above any other nominated contrast. Thus, for example, $R1$ is not above $RW11$ because $R1$ contains the mean of the W effect while $RW11$ contains the slope (but not the mean).

The situation is perhaps easier to understand if we use a less familiar notation. Write x_0 if we have taken a mean over x, and x_1 if we have taken a slope over x, and X or X_t for having done neither, without or with trimming. Then

$$R1 \rightarrow d_0\, r_1\, w_0$$

TABLE 9. Analysis-of-variance table after nomination and before aggregation. (Labels for all nominees (i.e., linear, linear-by-linear, etc.) are marked (n).)

Label	df	MS	DEN	F	Sig
$R(n)$	1	25426	94	270.5	0.01%
$D(n)$	1	348	94	3.70	10%
$W(n)$	1	4338	94	46.1	0.01%
$DR(n)$	1	36	94	.38	not
$RW(n)$	1	492	94	5.23	5%
$DW(n)$	1	3041	94	32.4	0.01%
$DRW(n)$	1	1089	94	11.6	0.5%
Rtrim	2	58	94	.62	not
Wtrim	5	59	94	.63	not
DRtrim	2	14	94	.15	not
RWtrim	17	49	94	.52	not
DWtrim	5	46	94	.49	not
DRWtrim	17	94	–		
(Total)	(55)	689			

$$RW11 \rightarrow d_0\ r_1\ w_1$$

while

$$R \rightarrow d_0\ R\ w_0$$

$$RW \rightarrow d_0\ R\ W$$

or

$$R_t = d_0\ R_t\ w_0$$

$$R_t\ W_t = d_0\ R_t\ W_t$$

so that the 3 basic questions are: Is w_0 above w_1? (No, but!) Is w_0 above W? (Yes!) Is w_0 above W_t? (Yes!) Where the parenthesized answers are for Interpretation One.

The issue is that a real (non-zero) slope *need not* — but is very likely to — imply a real (non-zero) mean. If the exact location of the mean for a factor is not any meaningful value, then the likelihood of a mean-free slope is small, so we may want to move away from Interpretation One.

It is far from clear when we ought to move all the way from Interpretation One to Interpretation Two and say "x_0 is over x_1". For the present, we recommend, in such circumstances, accepting Interpretation Two as a *possible alternative,* not an exclusive choice. One reason for this caution is the absence of any standardized way for x_1 to contribute to x_0 that is at all analogous to the standard contribution

$$\frac{\sigma^2}{\text{number of terms}}$$

of X to x_0.

Rules for Aggregation

Aggregation (as detailed by Green and Tukey) is the combination of lines in the analysis-of-variance table according to the values of their mean squares, using a rule of thumb (in philosophical contrast with significance testing, where non-significance is followed by pooling). The procedure is to start with the lowest remaining line (in terms of "above" with ties broken by value of mean square) and aggregate any other line with it which

a) is "above" the basic line and has mean square less than twice that of the basic line;

b) is NOT "above" any other line which does not satisfy (a).

A10. Aggregation in the Example

We start the aggregation with *DRW*trim, the lowest of the low, as we should. Since the line for each of the trimmed bouquets is above *DRW*trim and since each of the mean squares is less than (twice) that of *DRW*trim, the entire set of trimmed bouquets will be aggregated together. Additionally, since the nominated contrast *DR* 11 is above *DR*trim, it is also above *DRW*trim; and since the mean square of *DR* 11 is also less than (twice) that of *DRW*trim, *DR* 11 is also aggregated in with the trimmed bouquets. No other nominated contrasts have mean squares less than twice that of *DRW*trim, and so this step of aggregation ceases. We will identify the aggregated collection of all trimmed bouquets and *DR* 11 as "residual".

No other aggregations are possible, and so we are led to an aggregated analysis-of-variance table with 7 lines: 6 lines corresponding to the 6 largest nominated contrasts and a seventh line called residual with 49 degrees-of-freedom.

TABLE 10. Display ratios for the six largest nominated contrasts as members of 6-contrast and 5-contrast bouquets.

	6-Contrast Bouquet		After Super-Electing *R* 1 and Post-Trimming	
Contrast	Display Ratio	Working Value	Display Ratio	Working Value
R 1	98	(1.620)	237	(.674)
W 1	59	(1.119)	43	(1.534)
DW 11	68	(.804)	55	(1.009)
DRW 111	59	(.555)	49	(.674)
RW 11	65	(.336)	55	(.402)
D 1	141	(.132)	119	(.157)

We can choose to treat the 6 remaining nominated contrasts as a single bouquet of 6. In this case, we form the display ratios shown in the first two columns of Table 10. If we were "splitters," we might make a separate bouquet of $R1$, leaving the other 5 nominees in a single bouquet, which we will call "middle 5" (since the largest and smallest of the original 7 have been removed). The last two columns of Table 10 show the resulting display ratios.

The striking thing in Table 10, when compared with Table 6, is the proportionally large increase in the display ratio for $D1$. The suggestion, in view of the relative constancy of the display ratio of the next 4 contrasts, is that the size of the date main effect is larger than might be expected for the smallest of the unaggregated contrasts, possibly because part of one of the dates was "different".

The analysis-of-variance table after aggregation, collecting the 6 largest nominated contrasts into a bouquet and then electing (separating) out $R1$, has three lines, given in Table 11. The horizontalized plots for the three bouquets corresponding to Table 11 are shown in Figure 6. The essential difference between the first panel of Figure 6 and the second panel of Figure 4 is that we have now put the smallest nominated contrast, $DR11$, into the residual before constructing the nominated bouquet. (The size of contrast for $DR11$ is, in fact, exactly the median size of contrast of the 49 members of the residual poly-bouquet.) The main result of this exclusion of $DR11$ from the nominated bouquet is to inflate the size of the display ratio of the now-smallest member, $D1$.

TABLE 11. Analysis-of-variance table after nominating, aggregating, collecting all remaining nominated contrasts into a bouquet and electing out the largest.

Label	df	MS	DEN	F	Sig*
$R(n)$	1	25426	64	397	0.01%
middle 5	5	1862	64	29	0.01%
Residual	49	64			
Total	(55)	689			

* Notice that (a) significance level of F-values requires a large, rather unspecified multiplier for multiplicity and (b) 0.01% is the most extreme level considered.

Turning to the second panel of Figure 6, the horizontalized plot of the residual bouquet of 49, we see that our aggregation has eliminated the high values of display ratios for the smallest of the three-factor contrasts that we have become used to seeing. Instead, the smallest four contrasts within the residual bouquet (in order: *DW*13, *RW*33, *RW*24 and *W*6) have smaller display ratios than might be expected. This seems unlikely to mean anything. (The smallest three-factor contrasts, *DRW*133 and *DRW*116, are now the 8th and 11th smallest contrasts of the 49 and are in the bump seen on the left-hand side of the plot). The largest three contrasts are the three-factor contrasts *DRW*123, *DRW*112, *DRW*122, in decreasing order, each of which has a display ratio very slightly smaller than might be expected. In general, the display ratios of the contrasts in the residual bouquet are almost precisely what we would expect for random Gaussian noise with homogeneous variability.

Results after Aggregation

At the end of our aggregated analysis we have come to the following overall phenomenological picture:

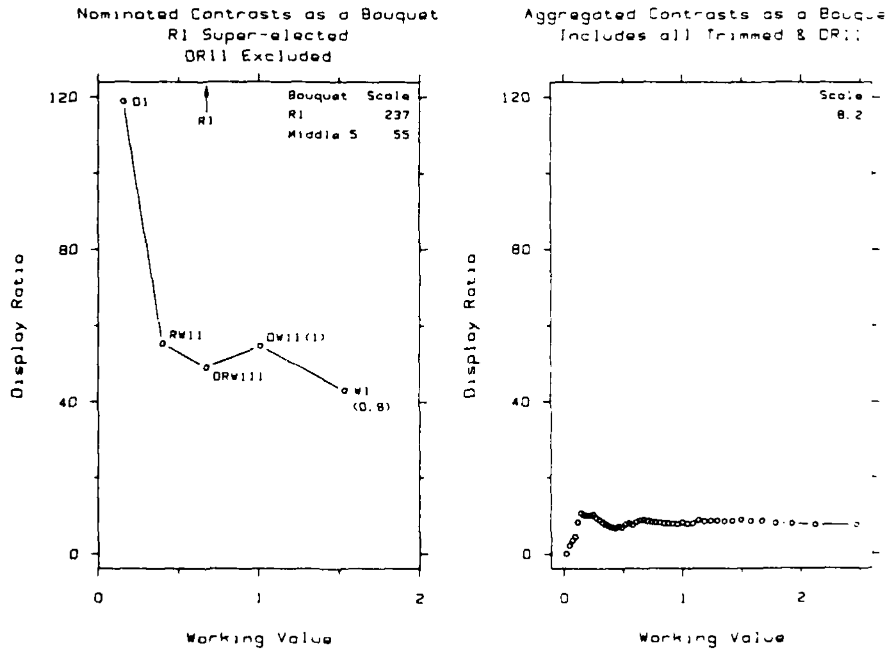


Figure 6. Person IB1 — D.L. (grams). Result of aggregation and election.

- large linear rate effect
- moderate linear-to-the-j contrasts for each of 5 other lines
- perhaps a few small exotics, calculation errors, etc. that remain unflagged
- a featureless "error body"

This is, of course, except for the small exotics, just what we saw when we didn't begin by aggregating, an encouraging agreement. All these choices of analysis can, and often should, lead to the same report (see section B5).

And under Interpretation Two?

Almost as an aside, we note that if we had considered the various nominated contrasts to be above each other, as might seem reasonable after aggregation (i.e., $R1$ is above $RW11$, $DR11$ and $DRW111$ but not above $DW11$), then the resulting aggregation would produce 4 lines rather than 3, namely:

- $R1$
- $D1$, $DW11$ and $W1$ (linear date and weight involving a mean on R)
- $DR11$, $RW11$ and $DRW11$ (linear date and weight involving a slope on R)
- all trimmed bouquets

This, too, gives an analysis worth looking at and thinking about.

PART B: DATA-GUIDED TRIMMING OF BOUQUETS

B1. Scales and Ratios-to-Scale

We are now used to plotting display ratios of the form

$$\frac{\text{size of contrast}}{\text{typical order statistic}}$$

in various ways. The general picture is that most ratios are estimating a residual variability (possibly differing from one bouquet to another), but that some, hopefully a few, are trying to reveal consistent contributions of some sort. To assess the size of the residual variability, it is thus natural to turn to the median (or conceivably the midmean) of these display ratios. It is this median that is shown as "scale" next to the bouquet labels in the horizontalized plots (Figures 2, 3, 4, 6). Because of the "dragging upward" phenomena we are not surprised to see this scale decrease somewhat as we trim the bouquets, removing opportunities for dragging upward.

Ratio-to-Scale

We have been judging the sizes of the display ratios both externally — looking across bouquets — and internally — within either the entire or a trimmed bouquet. For internal comparison, it is convenient to have numbers, and the natural number to look at is the

$$\text{ratio-to-scale} = \frac{\text{display ratio}}{\text{median display ratio}}.$$

(In other contexts, but not here, we may want to refer to this as an assassination or sterilizability ratio.) It is these ratios that are attached (in parentheses) to individual high points in the horizontalized plots (Figures 2, 3, 4, 6).

B2. Null Behavior of "Ratio-to-Scale"

It is helpful to know how large values of ratio-to-scale we are likely to see in a null situation, particularly how large a value we are likely to see for that contrast of largest size (in the bouquet at hand). The distribution of

$$\frac{\text{display ratio (for largest-size contrast)}}{\text{median display ratio (for all contrasts in the bouquet)}}$$

is easily simulated, starting with a sample of d "sizes" from a half-Gaussian. The resulting % points are given in Table 12.

Simulation Details

Each row of Table 12, corresponding to a bouquet of size d, was computed from the empirical distribution of the ratio-to-scale for the largest order statistic from a sample of size d from a half-Gaussian distribution. This empirical distribution was based on 2048 replicates. The half-Gaussian random variates were generated using the random normal generator of Kinderman and Monahan (1976), that generator using the McGill universal uniform generator (see Chambers, 1977), a combination of a 32-bit congruential with an independent 32-bit shift-register generator.

Simulation Error

To obtain an estimate of the variability of these simulated percent points, a method akin to balanced repeated replications (the multi-halving jackknife) was used. Based on the parity of the ith digit in the binary representation of the replication number (1-2048, in the order of generation), the collection of 2048 values of ratio-to-scale can be divided into 11 pairs of mutually orthogonal, interpenetrating

TABLE 12. Percent points from the distribution of the ratio-to-scale of the largest size of contrast for a sample of size *d* from the half-Gaussian (from simulations of 2048 replicates each – standard errors in parentheses).

Sample Size	Probability of Larger Value									
d	20%		10%		5%		1%		0.5%	
2	1.40	(.02)	1.67	(.01)	1.83	(.01)	1.95	(.01)	1.98	(.01)
3	1.22	(.03)	1.75	(.06)	2.48	(.15)	5.25	(.67)	6.71	(1.0)
4	1.30	(.03)	1.65	(.04)	2.12	(.08)	3.86	(.32)	4.80	(.40)
5	1.30	(.03)	1.70	(.05)	2.14	(.06)	3.86	(.25)	4.61	(.43)
6	1.30	(.01)	1.60	(.04)	1.95	(.04)	3.22	(.18)	3.90	(.27)
7	1.28	(.02)	1.56	(.03)	1.91	(.06)	2.98	(.12)	3.70	(.30)
8	1.28	(.02)	1.53	(.02)	1.80	(.05)	2.64	(.08)	3.19	(.24)
9	1.24	(.01)	1.53	(.02)	1.80	(.04)	2.58	(.11)	3.01	(.17)
10	1.27	(.03)	1.51	(.03)	1.78	(.04)	2.61	(.09)	2.88	(.10)
15	1.24	(.02)	1.46	(.02)	1.65	(.02)	2.27	(.11)	2.54	(.13)
20	1.22	(.01)	1.39	(.01)	1.55	(.02)	2.01	(.03)	2.17	(.03)
30	1.18	(.01)	1.30	(.02)	1.44	(.02)	1.74	(.04)	1.88	(.05)
approx. (for $d > 2$)	$1.20 + \frac{.36}{d}$		$1.30 + \frac{1.4}{d}$		$1.38 + \frac{3.24}{d}$		$1.42 + \frac{10.7}{d}$		$1.43 + \frac{15}{d}$	

half-samples. From each pair of half-samples, two estimates (one for each half-sample) of the percentage points can be obtained, and an estimate of the variance of the (full-sample) percent point, worth perhaps 1 df, comes from ¼ the squared difference of the half-sample estimates. The standard errors, shown in parentheses in Table 12, are the square roots of the average of the 11 separate variance estimates and each is worth (optimistically) perhaps 11 df.

d = 2 is Special

The anomalous appearance of the percent points for $d = 2$, relative to the general pattern exhibited for the larger bouquet sizes, is due to the special constraint placed on the maximum size of the ratio-to-scale for a sample of size 2. Since we have defined the denominator of the ratio-to-scale as the median of the two display ratios, the values of the ratio-to-scale for a sample of size 2 are bounded above by 2.

Some Approximations

Approximations for the values of these percent points, valid for $d > 2$, are given at the bottom of Table 12. These approximations were derived by fitting a linear dependence of the percent point on $1/d$, a column at a time, by a simple resistant regression. To preserve monotonicity between columns for large values of d, the last two approximations were modified by -.02 and +.02, respectively. The approximations have relative errors of less than 8% throughout the pertinent ($d \geqslant 3$) entries of the table.

In terms of the estimated standard errors, 30 of the 55 approximations were within 1 standard error of the simulated percentage point, and 44 of the 55 approximations were within 2 standard errors. (The null comparison calls for 36 and 51, respectively.) In these terms, the approximations were relatively better for the columns for 5%, 1% and 0.5% probabilities of a larger value (P), where 31 of the 33 approximations were within two standard errors.

The maximum absolute estimated error for the approximations is, by column, .10 for $P = 20\%$, .12 for $P = 10\%$, .11 for $P = 5\%$, .30 for $P = 1\%$ and .36 for $P = 0.5\%$.

B3. Post-Trimmed Bouquets; Election

The value of the ratio-to-scale of the largest size of contrast within a bouquet (which may be either already pretrimmed or an original bouquet) provides us with a criterion for use in assessing the amount of attention that should be paid to that particular contrast in terms of describing the data. The largest size of contrast within the bouquet can be considered as indicating an interesting, potentially credible, component of the total information collectively imparted by the bouquet if the value of its ratio-to-scale is larger than some selected threshold value, where the threshold value can be selected in view of our Table 12 and probably should depend on the number of contrasts in the bouquet.

Having found such a display ratio, we can *elect* it for special attention and post-trim the bouquet, creating two new bouquets: one consisting solely of the elected contrast — designated as potentially interesting -- and the other, post-trimmed bouquet consisting of the remainder. We then proceed as we did for nomination and pre-trimming, recomputing the display ratios, standing ready to assess the relative importance of the elected contrast in the experiment as a whole in terms of its relative standing in terms of display ratio.

We can, and may need to, repeat the election process before doing anything else, comparing the ratio-to-scale of the (now) largest size of contrast within the post-trimmed bouquet with an appropriate threshold value and proceeding as above if the threshold value is exceeded. (Experience will show, we believe, to what extent such recursive calculation is wise.) At some stage, which might very well be the assessment of the largest size of contrast within the initial (not post-trimmed) bouquet, the ratio-to-scale of the largest size of contrast within the current bouquet will not exceed the threshold value. The interpretation is that this contrast, and hence all remaining contrasts of smaller size, does not appear individually to be capturing a significant amount of the information embodied in the bouquet beyond that expected by the simple (null) half-Gaussian model. At this stage, post-trimming certainly ceases.

We will end up splitting the initial bouquet into two sets of contrasts: the first set, which may be empty, consists of the elected contrasts, the largest in the initial bouquet, each of which individually appears to account for an important amount of the information embodied in the bouquet. The second set is the post-trimmed bouquet and consists of the remaining (smaller) contrasts, which are deemed not to be *separately* providing significant information. We return later to the possibility of extracting additional information from the post-trimmed bouquet.

Treating the Elected Contrasts

The elected contrasts can be treated either individually, source bouquet by source bouquet, or as members of a single bouquet, the elected bouquet, analogously to our treatment of the nominated contrasts. They could even be combined together with the nominated contrasts into a single bouquet, the nominated-plus-elected bouquet. In considering the elected contrasts, some attention should be paid to issues of multiplicity.

We can apply the post-trimming procedure to the nominated bouquet itself (as well as to the elected bouquet and the nominated-plus-elected bouquet). Contrasts flagged as too large in terms of their ratio-to-scale within such a bouquet will be referred to as being *super-elected,* since they have distinguished themselves above the other contrasts, already distinguished by nomination or election.

B4. Election (Post-Trimming) in the Example

Although we have advocated pretrimming (nominating) the $D1$, $R1$, $W1$, $DR11$, $DW11$, $RW11$ and $DRW111$ linear-to-the-j contrasts as a general maxim, it is interesting to consider what would have been the effect if instead we had retained the full bouquets and elected large contrasts exceeding some threshold value(s), probably less than or equal to 2.5. Returning to Figure 2, we have $R1$, $W1$ and $DW11$ as the three most obvious candidates for election, having ratios-to-scale of 7.8, 4.4 and 3.4, respectively, each of which is above the corresponding 1% point for the appropriate bouquet size from Table 12. These, of course, had the highest display ratios after pretrimming, substantially exceeding the levels of all other contrasts. A horizontalized plot of the result of electing $R1$, $W1$ and $DW11$, post-trimming the R, W and DW bouquets and leaving the other bouquets alone, is shown in Figure 7.

The vertical axis of the plot has been truncated at 50 to show detail, and so the display ratios of the elected contrasts, which we are treating as individual bouquets, are off scale. These ratios are the same as in Table 5 and also correspond to the values of "scale" shown in the inserts of the horizontalized plot. This plot is, of course, a middle ground between the horizontalized plot of the original bouquets (Figure 2) and the horizontalized plot of the pretrimmed bouquets (Figure 3). Depending on the values of threshold selected, we could have also considered electing $DRW111$, $RW11$ and $W2$, in that order. Since the threshold values for suggesting this do not reach the 10% points of Table 12 in each case, we shall not do so, but instead will turn to Figure 3.

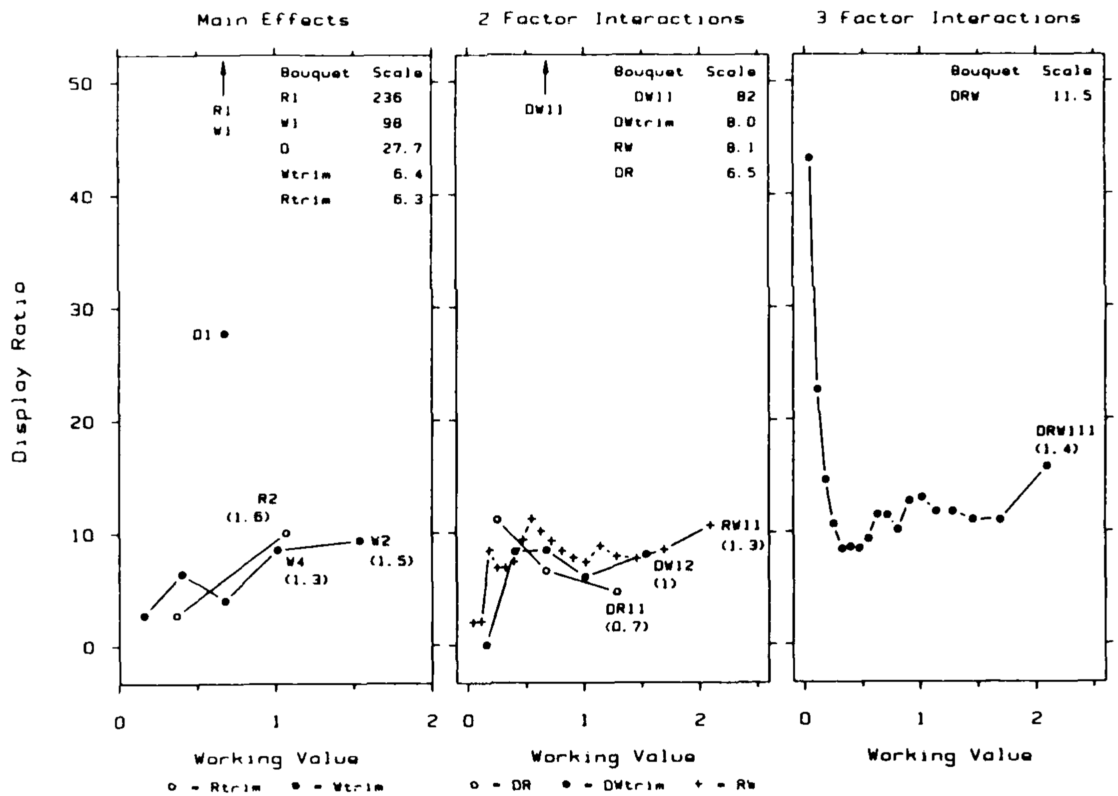


Figure 7. Person IB1 — D.L. in grams — polynomial contrasts . Post-trimmed bouquets — R1, W1 and DW11 elected.

On looking at the values of ratio-to-scale for the largest contrasts in the pretrimmed bouquets in Figure 3, we see that $R2$ and $W2$ may be of marginal interest and that no other contrasts distinguish themselves.

Clearly, a large part of the story is being told by the six largest nominated contrasts: $R1$, $W1$, $DW11$, $DRW111$, $RW11$ and $D1$, in apparent order of importance. Of these, we have already super-elected $R1$ as the single contrast displaying the most information about the experiment.

It is possible to push the analysis of the responses for person IB1 further, but in order to do so more readily, and to demonstrate more clearly certain other characteristics of our horizontalized plots, we will reformulate the response. First, however, we summarize what we have found so far about IB1's responses.

B5. The Final Outcomes for the First 56 Numbers

The previous sections have indicated that, as far as difference limen in grams is concerned, the relationship between the responses of IB1 and the various factors is largely captured by the nominated contrasts $R1$, $W1$, $DW11$, $DRW111$, $RW11$, $D1$ and $DR11$. Let us model, with i = date, j = rate and k = (initial) weight, the response in terms of the linear-to-the-j contrasts as

$$y_{ijk} = a_0 + b_D d(i) + b_R r(j) + b_W w(k) +$$

$$+ b_{DW} d(i) w(k) + b_{RW} r(j) w(k) + b_{DR} d(i) r(j) + b_{DRW} d(i) r(j) w(k) + z_{ijk}$$

where, for convenient comparison of coefficients, we have taken

$$\Sigma d(i) = \Sigma r(j) = \Sigma w(k) = 0$$

$$\Sigma d(i)^2 = 1$$

$$\Sigma r(j)^2 = 3$$

$$\Sigma w(k)^2 = 6 .$$

Then we can summarize our results by

$$a_0 = \text{centercept} = 50.4$$

$$b_R = \text{rate slope} = 24.6$$

and by the other linear-to-the-h ($h = 1, 2, 3$) contrasts in the $2 \times 2 \times 2$ table:

	No D		D	
	No R	R	No R	R
No W	a_o	b_R	$b_D = 3.5$	$b_{DR} = 1.3$
W	$b_W = -9.5$	$b_{RW} = -3.7$	$b_{DW} = -11.3$	$b_{DRW} = -7.8$

The residuals are

Rate (gm/30 sec)	Date	Initial Weight (Grams) 100	150	200	250	300	350	400
50	1	−1.3	1.5	3.0	−2.8	2.0	2.3	1.9
	2	8.5	0.3	2.1	0.5	0.8	0.8	−0.4
100	1	6.2	0.4	−8.4	−8.4	2.2	0.0	−0.1
	2	−3.2	3.8	−0.8	−4.9	−5.9	−1.8	3.5
150	1	2.6	−5.9	0.7	3.5	0.4	1.6	−6.4
	2	−16.2	−2.4	−3.4	1.6	−10.9	9.8	2.7
200	1	−4.8	1.8	7.2	0.9	−4.3	0.9	3.4
	2	26.6	14.5	−22.1	−20.7	8.2	−0.5	9.7

The various horizontalized plots show us that, so far as orthogonal polynomial contrasts go — simple or multiple — there is no appreciable evidence of needing a more detailed description.

We are coming out as we should; using numbers to describe our results and pictures to show no need to go further in the terms we have considered.

PART C: ANALYZING IB1's PERFORMANCE IN OTHER TERMS

C1. Reformulating the Response

It is often the case that the original response variable, the values of which were recorded (or calculated) in the process of the experiment, is not the best variable to use in the analysis. It is sometimes possible to find a reformulation of the response data which yields a simpler, clearer set of relationships between the dependent variable and the factors of the experiment. This is achieved when there are fewer important interactions and when the important main effects become more pronounced in reference to the background level. This can sometimes be achieved by a more trenchant change in the definition of the response which involves the values of one or more factors as well as the response, one that may largely remove the impact of a previously important main effect.

Besides seeking to simplify the relationships necessary to adequately describe the data, we shall be delighted if we can also make the variability of the response variable approximately homogeneous.

In considering the results of our initial analysis of the response data (as D.L. in grams) for our example person IB1, we noticed that the linear contrast for the rate main effect, $R1$, was far and away the most important single contrast in the experiment. A look at Figure 1 suggests that the relationship between difference limen in grams and rate might be close to being a proportional one; Table 13 confirms this. In this table we show the average of the difference limen values (in grams) for IB1 for each level of rate, averaging across all levels of data by weight within each rate level. The second line of the table is

TABLE 13. Average difference limen in grams by rate and its ratio to rate for person IB1.

	Rate (grams/30 seconds)			
	50	100	150	200
Average D.L. (in grams)	23.2	39.7	58.4	80.5
$30(\frac{\text{average D.L}}{\text{rate}})$ (in seconds)	13.9	11.9	11.7	12.1

the ratio of this average level in grams to the rate. To produce simple units, since rate is measured in grams per 30 seconds, we multiply by 30 so that the response is now in seconds of time. Since the original response variable, D.L., was the number of grams of water added to a pail at a constant rate until the person could detect a difference in pull, the new response (given in the second line of Table 13) is a response time.

The relatively constant response time (relatively constant compared to a change from 23 grams to 80 grams) for the various levels of rate implies that the relationship between response and rate can be largely explained by assuming that the person responds after a constant time, regardless of the rate. (This was found by Green and Tukey to hold in a collective analysis of all 8 persons.)

Since the large rate effect can be substantially explained in the above manner, we can obtain a simpler analysis by changing our response from difference limen (in grams) to

$$\text{response time} = \frac{\text{difference limen}}{\text{rate}} \times 30 \text{ (in seconds)}.$$

Re-Expressing the Reformulated Response

We shall, in fact, go slightly farther. In their analyses of the entire experiment, Green and Tukey found, when the dependent variable was re-expressed as response time, that the standard deviations were approximately linearly related to the means. This can be seen most clearly by comparing persons. However, a plot of the residuals versus predicted values for a linear-to-the-j fit of the response time for our example person also shows an increase in the variability as the predicted response time increases. In order to produce more homogeneous variability, Green and Tukey re-expressed the reformulated response on the log scale. We will do the same, and so our new response is

log(response time),

the natural log of the response time as defined above.

Figure 8 shows the relationship between our new response variable and rate within each combination of weight and date levels. The actual data values appear in Table 14.

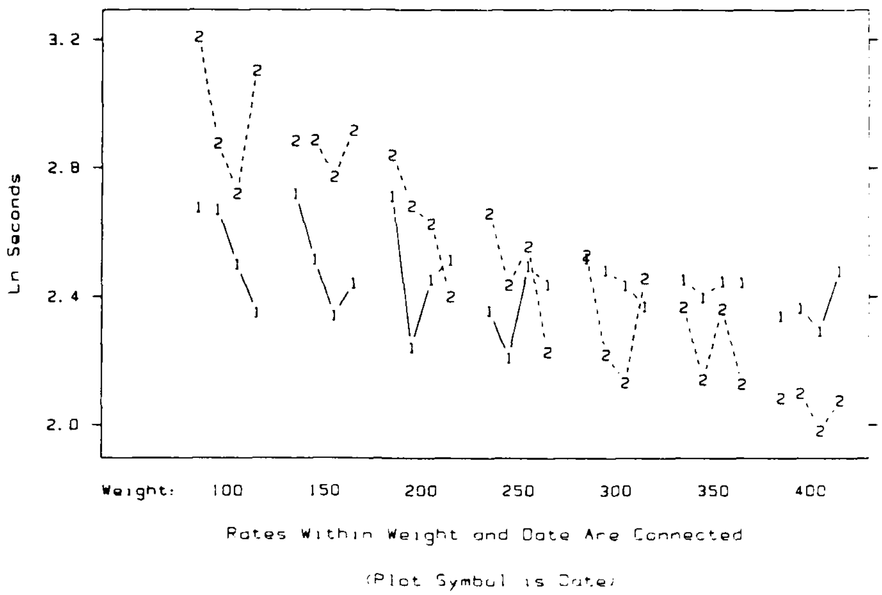


Figure 8. Person IB1: male, blind. Log response time (Ln seconds).

TABLE 14. Log response time in log seconds for person IB1.

Rate (gm/30 sec)	Date	Initial Weight (grams) 100	150	200	250	300	350	400
50	1	2.68	2.72	2.71	2.36	2.52	2.45	2.34
	2	3.21	2.88	2.84	2.66	2.53	2.37	2.08
100	1	2.67	2.51	2.24	2.21	2.48	2.40	2.37
	2	2.88	2.89	2.68	2.44	2.22	2.14	2.10
150	1	2.50	2.34	2.45	2.49	2.44	2.45	2.29
	2	2.72	2.77	2.63	2.56	2.13	2.36	1.98
200	1	2.35	2.44	2.51	2.44	2.37	2.45	2.48
	2	3.10	2.92	2.40	2.23	2.46	2.13	2.08

C2. A First Analysis of Log Response Time

Having reformulated our response from D.L. (in grams) to log response time (in log seconds), we proceed with an analysis of the apparent relationships involving the various factors (data, rate and weight, as before) via the horizontalized plotting techniques. As an initial step, we will use the orthogonal polynomials, as before, to produce the single-degree-of-freedom contrasts which define our various bouquets. As in the analysis of D.L. in grams, we nominate the linear-to-the-j contrasts $D1$, $R1$, $W1$, $DR11$, $DW11$, $RW11$ and $DRW111$ as *a priori* important and pretrim our bouquets. Going through the same steps as in the analysis for D.L., we reach the plots of Figure 9. (We note, in passing, that $RW11$ and $DRW111$ are now *not* the largest contrasts in their original bouquets — we had already nominated them anyway.)

In our plots of the display ratios (in units of log seconds) in Figure 9, we have truncated the vertical axis at 0.6 in order to better show the detail. This truncation has put the two largest display ratios, $W1$ and $DW11$, off scale. The display ratios for the nominated contrasts, which we are treating as 7 separate bouquets, are listed in Table 15. Also included in Table 15, for comparison purposes, are the display ratios (in units of grams) from the pretrimmed analysis of the original response (in grams), as well as the median display ratio across all 55 contrasts for both responses and the ratio of display ratio to these medians.

TABLE 15. Display ratios (in units of $\log_e$ seconds) for the polynomial analysis of log response time for person IB1 (display ratios for D.L. in grams included for comparison).

	Log Response Time		D.L. In Grams	
Contrast	Display Ratio ($\log_e$ Seconds)	Ratio To Median	Display Ratio (Grams)	Ratio To Median
$W1$	1.97	16.4	98	10.9
$DW11$	1.42	11.8	82	9.1
$R1$	.54	4.5	237	26.3
$D1$	.34	2.8	28	3.1
$RW11$	.33	2.8	33	3.7
$DRW111$	.24	2.0	49	5.4
$DR11$	.16	1.3	9	1.0
Median Display Ratio Over All 55 Contrasts	.12		9	

Comparison of Analyses

We can see that the relative importance of the *R* 1 contrast has been considerably reduced by the reformulation but that, in comparison to the median display-ratio, it still merits appreciable attention. The relative apparent importance of the weight and date-by-weight linear contrasts remains high and has, in fact, been increased by going to log response time. The relative importance of the remaining four nominated contrasts has generally been decreased.

Table 16 shows the display ratios and ratios-to-scale for the nominated contrasts when they are treated as the 7-contrast nominated bouquet. Based on the sizes of their ratios-to-scale, we might conceivably consider super-electing *W* 1 and *DW* 11 as the most important contrasts, but their relative importance over the remainder of the nominated contrasts seems somewhat slight.

Examining the ratios-to-scale for the largest sizes of contrasts within the trimmed bouquets, we find these to be generally small, the largest being 1.4 for *RW* 32, the cubic-by-quadratic interaction for rate-by-weight, and also for *DW* 16, the linear-by-6th-degree interaction for date-by-weight. Because of the relatively small sizes of these ratios-to-scale and because of the unpromising nature of the contrasts associated with them, we will not elect either one.

General Levels

In looking at the general level of the display ratios for the trimmed bouquets (also reflected by their value of "scale" indicated on the plots), we feel that for the most part these display ratios are measuring noise. There are two exceptions — the first being the relatively large value of the display ratios of the smallest size of contrast within the three-factor bouquet — which we again take as measuring some type of isolated error or granularity and accordingly ignore. The other exception corresponds to the trimmed rate bouquet, measuring the quadratic and cubic contributions of rate, and is of more interest. The levels of the display ratios for both the contrasts in this bouquet are nearly the same and are noticeably above the general level ("scale") of the other trimmed bouquets and above the levels of two of the nominated contrasts. This type of general inflation of all levels of a trimmed bouquet, with no contrast indicated as individually important, could be indicative of a particular phenomenon, the discussion of which we turn to next.

TABLE 16. Display ratios (in units of $\log_e$ seconds) for the nominated contrasts, treating them as a 7-contrast nominated bouquet.

	Log Response Time For Person IB1		
Contrast	Display Ratio	Working Value	Ratio-to-Scale
W 1	.79	(1.691)	1.42
DW 11	.80	(1.208)	1.43
R 1	.41	(.908)	.73
D 1	.34	(.674)	.61
RW 11	.48	(.472)	.86
DRW 11	.55	(.288)	1.00
DR 11	.97	(.114)	1.74
	scale = .55		

PART D: RETHINKING A SCISSION INTO CONTRASTS

D1. Spreading of Contributions across Contrasts

We have previously stated that each trimmed bouquet, at the end of post-trimming, where the ratios-to-scale of the remaining contrasts are not sufficiently large, is deemed to consist of contrasts which *individually* are not imparting significant information about the relationships between the response variable and the collective components of the factor(s) embodied in the (trimmed) bouquet. This need not mean that the trimmed bouquet is certain not to be telling us anything of importance about the data.

The elected contrasts from post-trimming are each, potentially, individually representing systematic relationships in the data (of varying degrees of strength depending on their sizes of trimmed-out display ratios). It is possible that there are other, real, systematic relationships within the data which are not individually captured by any of the various contrasts which have been selected to define the (original untrimmed) bouquets. Such a systematic relationship is then jointly indicated by a number of contrasts, and its actual size is spread across those contrasts.

The effect of such a situation might be a bouquet (trimmed or untrimmed) where the general level of the display ratios (the scale) is at the background noise level and where the largest few sizes of contrast correspond to the contrasts jointly indicating the systematic relationship. Because this relationship is spread among a number of the contrasts, it is possible that not all, or even none, of these contrasts will be flagged in post-trimming as individually indicating an interesting contribution.

If the systematic relationship is spread across enough of the defining contrasts for a bouquet, the scale for that bouquet will be inflated, and the plot of the display ratios for the bouquet will be at a general level above the background noise level, although no individual contrasts in the bouquet may be flagged as individually interesting.

It is sometimes possible — either in these two circumstances or, better, initially — to select a different bouquet of defining contrasts for the line of the analysis-of-variance table which will more nearly isolate the systematic contributions, each into a single (different) contrast, and result in a potentially simpler account of the interrelationships in the data.

D2. Some Useful Bouquets of Contrasts

Some interesting bouquets of orthogonal contrasts emphasize the ordering of the versions of a factor. A classical example of those using only order are the Helmert contrasts, which, for example, may be formulated to compare the response value of the first version with the average of the remaining versions, the value of the second with the average of all but the first two, and so on, the last such contrast comparing the response value of the next-to-last version with that of the last version. We will call these Helmert SFP contrasts, for "Starting with First Point." There are also Helmert SLP contrasts, starting with the last point. In our situation, however, we would like to use both order *and* value. In particular, we will often want to include *the* linear contrast.

LPO's and FPO's

An interesting alternative was recently considered by Daniel (1985), who notes that if a set of m responses at equally spaced versions of a factor are nearly linear in that factor, then a commonly observed deviation from linearity is localized at one end, the remaining $m-1$ points falling close to a straight line. He defines the contrast LPO_m, for Last Point Off of m, which measures the deviation of the mth point in the sequence from its predicted value based on a least squares line through the previous $m-1$ points. Table 17 shows these Last Point Off contrasts for $m = 3, ..., 7$. Given n equispaced versions of a factor, the collection $\{LPO_n, LPO_{n-1}, ..., LPO_3, L\}$ defines a set of orthogonal contrasts, where L is the ordinary linear contrast and LPO_i compares the observed value at the ith level with the predicted value from the line through the values for the first $i-1$ levels.

General coefficients for LPO_m, $m \geqslant 3$:

$$i\text{th value} = m + 1 - 3i,\ i < m$$

$$\text{end } (m\text{th value}) = \tfrac{1}{2}(m-1)(m-2)$$

$$\text{Sum of Squares} = \tfrac{1}{4}(m-2)(m-1)m(m+1)\ .$$

If we use the last column, labeled "(j)", in Table 17 as our ordering, we get Daniel's First Point Off contrasts. Ordinarily, in those circumstances where *LPO* or *FPO* contrasts are likely to be helpful, either advance insight or data behavior will make it clear which to select. But there may be doubt.

TABLE 17. Last Point Off contrasts for equally spaced levels illustrated for m from 3 to 7 (* marks special point).

i	LPO_3	LPO_4	LPO_5	LPO_6	LPO_7	(j)
1	1	2	3	4	5	7
2	−2	−1	0	1	2	6
3	1*	−4	−3	−2	−1	5
4		3*	−6	−5	−4	4
5			6*	−8	−7	3
6				10*	−10	2
7					15*	1
SSq	6	30	90	210	420	

EPO's

In such doubtful cases, the *EPO* contrasts, for End Point Off, which treat both ends more nearly symmetrically, may be in order. These can be easily made up from *FPO*'s and *LPO*'s. Table 18 shows examples for $n = 7$ and $n = 6$, where the slight "preference" has been given to *LPO*'s. (In the world as a whole, we believe there is at least as much curvature near the upper end of the range as near the lower end.)

TABLE 18. Double-ended (EPO) contrasts for equally spaced levels for $n = 7$ and $n = 6$ (* marks special point – note slight preference for LPO).

$n = 7$ Rank	L	$m = 7$	$m = 6$	$m = 5$	$m = 4$	$m = 3$
1 of 7	−3	5	10*	–	–	–
2 of 7	−2	2	−8	3	3*	–
3 of 7	−1	−1	−5	0	−4	1
4 of 7	0	−4	−2	−3	−1	−2
5 of 7	1	−7	1	−6	2	1*
6 of 7	2	−10	4	6*	–	–
7 of 7	3	15*	–	–	–	–
SSq	28	420	210	90	30	6
Identity	L	LPO_7	FPO_6	LPO_5	FPO_4	LPO_3

$n = 6$ Rank	L	$m = 6$	$m = 5$	$m = 4$	$m = 3$
1 of 6	−5	4	6*	–	–
2 of 6	−3	1	−6	2	1*
3 of 6	−1	−2	−3	−1	−2
4 of 6	1	−5	0	−4	1
5 of 6	3	−8	3	3*	–
6 of 6	5	10*	–	–	–
SSq	70	210	90	30	6
Identity	L	LPO_6	FPO_5	LPO_4	FPO_3

SEPO's

If it is really important to have symmetry, to the extent that we do not mind irrational coefficients, we can arrange for it by combining pairs of *LPO*'s and *FPO*'s. We will call these *SEPO* contrasts, for Symmetric End Point Off. For $n = 7$, for example, we can use L, $FPO_7 + FPO_6\sqrt{2}$, $LPO_7 + LPO_6\sqrt{2}$, $FPO_5 + FPO_4\sqrt{3}$, $LPO_5 + LPO_4\sqrt{3}$, $LPO_3 - FPO_3$. In these contrasts, both FPO_7 and FPO_6 treat the first data point (of the 7) as special, LPO_7 and LPO_6 the last data point, FPO_5 and FPO_4 treat the second data point (of the original 7) as special and compare it with the next 4 and 3 points respectively, and so on. For either initial or intermediate values of $m \geqslant 4$, the First Point *SEPO* combination is $FPO_m + FPO_{m-1}\left(\frac{m+1}{m-3}\right)^{1/2}$, which has sum of squared coefficients $\frac{m(m+1)(m-2)}{2}[(m-1) + \sqrt{(m+1)(m-3)}\,]$.

Double-Ended Helmert Contrasts

We can also define a double-ended set of Helmert-type contrasts, as illustrated in Table 19.

General coefficients for m non-zero entries ($m \geqslant 3$):

$$\text{end values} = \frac{m-2}{2} \pm \frac{\sqrt{m(m-2)}}{2}$$

$$\text{inner values} = -1,$$

$$\text{Sum of Squares} = m(m-2).$$

D3. Alternative Descriptions of the (Log-Response) Rate Effect

We have noted that the trimmed rate bouquet (consisting of $R3$ and $R2$) has display ratios which are nearly equal and which appear high relative to the assumed background level. From Figure 9 we can see that the general level for the trimmed rate bouquet, as measured by its scale, .285, is 3 times that of the apparent background level (.092) as measured by the median scale of the 6 trimmed bouquets. A pattern of display ratios such as this, given its high level relative to the background, is suggestive of the spreading of a systematic relationship across contrasts.

TABLE 19. Double-ended Helmert-type contrasts for equally spaced levels illustrated for $n = 4$ and $n = 7$ (* marks special point).

$n = 4$	$m = 4$	$m = 2$	$m = 4$
1 of 4	2.414*		−.414
2 of 4	−1	+1*	−1
3 of 4	−1	−1	−1
4 of 4	−.414		2.414*
SSq	8	2	8

$n = 7$	$m = 7$	$m = 5$	$m = 3$	$m = 3$	$m = 5$	$m = 7$
1 of 7	5.458*					−.458
2 of 7	−1	3.436*			−.436	−1
3 of 7	−1	−1	1.366*	−.366	−1	−1
4 of 7	−1	−1	−1	−1	−1	−1
5 of 7	−1	−1	−.366	1.366*	−1	−1
6 of 7	−1	−4.36			3.436*	−1
7 of 7	−.458					5.458*
SSq	35	15	3	3	15	35

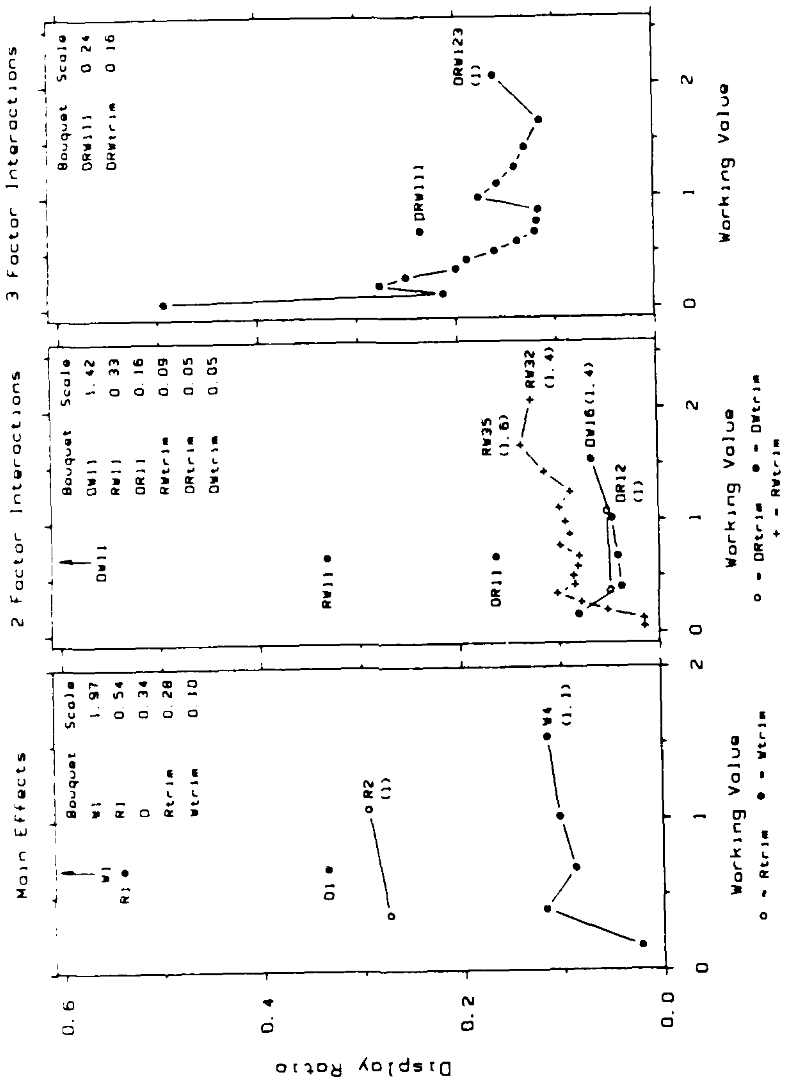


Figure 9. Person IB1 — log response time in Ln seconds. Polynomial contrasts — pretrimmed bouquets.

In fact, looking at the values of the size of contrast for the polynomial decomposition of the rate main effect, given in Table 20, we see that the linear and quadratic contrasts are both relatively large and of roughly the same magnitude.

Since the rate main effect has only three degrees of freedom, this near equality of the two largest sizes-of-contrast suggests that another orthogonal decomposition might produce a simpler description of the relationship between log response time and rate. To help understand if this is possible, we consider Figure 10, the plot of average values of log response time by rate for each date (averaging across weight within each combination of rate and date). The plot firstly shows a similar relationship between average log response time and rate within date, with a minor difference in the slopes of the simple linear fits of log response time vs rate — accounting for the moderate *DR* 11 display ratio. More importantly, within a given date the levels of the response variable for the latter three rates (100, 150 and 200) are all at roughly the same value and notably lower than the level of log response time for the rate of 50 grams/30 seconds.

Results for Various Bouquets

This latter observation suggests that a different bouquet of orthogonal contrasts might usefully be considered in defining the rate effect. Specifically, we want a set of contrasts which emphasize the ordering of the levels of the factor, such as those given in the last section.

Table 21 shows the results of applying a number of different bouquets of contrasts to the average values of the log response time

TABLE 20. Values of the size of contrast for the polynomial decomposition of the rate main effect for IB1 (log response time).

Contrast	Size of Contrast ($\log_e$ Seconds)
R 1	.365
R 2	.315
R 3	.100

for each of the rates. These averages are shown as the first line of the table. The remainder of the table consists of the sizes of contrast and display ratios that would be obtained if each of the bouquets given in the exhibit were used in defining the rate main effect. The sizes of contrast are those which would be obtained from a full analysis (and are $\sqrt{14}$ times the normalized values which would be obtained by applying the contrasts to the averages by rate). The display ratios are for the original three-contrast bouquet (with working values 1.282, .674, and .253). We have arranged the terms in each set of contrasts so that the resulting sizes of contrast come out in descending order. (Notice that *EPO* and *LPO* are identical for $n = 4$.) Clearly, in view of the display ratios, if spreading is occurring with the polynomial bouquet, then it is also occurring with the *EPO, LPO, FPO,* and *SEPO* bouquets of contrasts. The common element is, of course, the inclusion of the linear contrast.

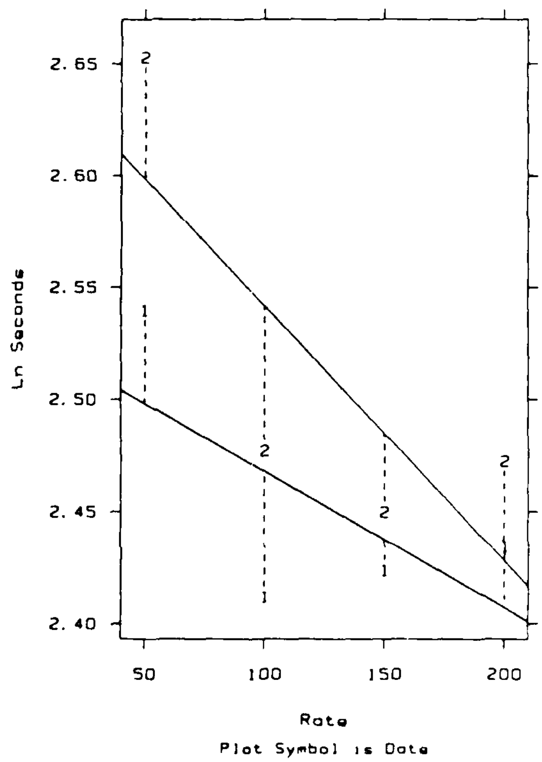


Figure 10. Person IB1 log response time vs rate, by date (lines are the linear fit).

TABLE 21. Values of size of contrast and display ratio for various bouquets of orthogonal decompositions of rate effects (person IB1 — log response time).

		Rate 50	100	150	200		
Average Response By Rate		2.596	2.445	2.437	2.454		
Bouquet						Size of Contrast	Display Ratio
Polynomial	L	−3	−1	1	3	.365	.284
	Q	1	−1	−1	1	.315	.470
	C	−1	3	−3	1	.100	.400
$EPO(=LPO)$	L	−3	−1	1	3	.365	.284
	LPO_4	2	−1	−4	3*	.250	.373
	FPO_3	1*	−2	1	0	.220	.880
FPO	L	−3	−1	1	3	.365	.284
	FPO_4	3*	−4	−1	2	.329	.491
	FPO_3	0	1*	−2	1	.037	.148
$SEPO$	L	−3	−1	1	3	.365	.284
	$FPO_4 + \sqrt{5}FPO_3$	5.236*	−8.472	1.236	2	.294	.439
	$LPO_4 + \sqrt{5}LPO_3$	2	1.236	−8.472	5.236*	.152	.608
Helmert	50 vs rest	3*	−1	−1	−1	.490	.382
SFP	150 vs 200	0	0	1*	−1	.044	.065
	100 vs 150, 200	0	2*	−1	−1	.002	.009
Double-ended	50 vs rest	2.414*	−1	−1	−.414	.490	.383
Helmert	200 vs rest	−.414	−1	−1	2.414*	.044	.066
	100 vs 150	0	1*	−1	0	.020	.081

* Marks special point, contrasts arranged in order of size of contrast, display ratios are for bouquets of 3 contrasts.

If we consider the display ratios for the Helmert *SFP* bouquet, we see that the majority of the information about the effect of rate on log response time is captured in the first contrast, the one comparing the response at rate 50 with the average of the other responses. While it is difficult to justify nominating this contrast, we can assuredly elect it, as its ratio-to-scale is 5.7. (According to Table 12, 5.7 is beyond the 0.5% level. Thus if we make an allowance for multiplicity of between 6 — the number of alternative bouquets in Table 21 — and 8 — the largest number of alternative bouquets we might reasonably have considered, we are still well beyond 5%.) The double-ended Helmert contrasts also produce much the same result.

D4. The Example after Rescission

Adopting the Helmert *SFP* contrasts as our scission of rate, and appropriately adjusting the definitions of all two- and three-factor contrasts which involve rate, will produce the horizontalized plots shown in Figure 11.

In the plots we use the following notation: $r1$ is the Helmert *SFP* contrast comparing the first rate (i.e., 50 gms/30 seconds) with the average of the remainder, $r2$ compares the second rate with the average of 3rd and 4th, and $r3$ compares the third rate with the fourth. The two-factor interactions involving rate are obtained as the outer product of the Helmert contrasts for rate and the polynomial contrasts for the other factor. Thus, $rW11$ is the interaction involving the first Helmert contrast for rate and the linear polynomial contrast for weight. Similarly, the factor contrast, $DrW123$, combines the day-to-day difference, the second Helmert contrast for rate, and the cubic-polynomial contrast for weight.

As we have been doing all along, we have nominated the linear-to-the-j contrasts which *do not* involve rate (i.e., $D1$, $W1$ and $DW11$) as *a priori* potentially important contrasts. As mentioned above, it is difficult to justify nomination of any of the Helmert contrasts, which is why no contrasts involving rate have been nominated. We have, however, elected $r1$ as an *a posteriori* important contrast in view of its ratio-to-scale within the full three-contrast Helmert *SFP* bouquet. No other contrasts involving rate can be elected for any reasonable threshold value.

Using the Helmert contrasts as our scission of rate has produced a simplification in the apparent relationship between the response and the three factors. This rescission has eliminated the apparent importance of any interaction involving rate in an adequate

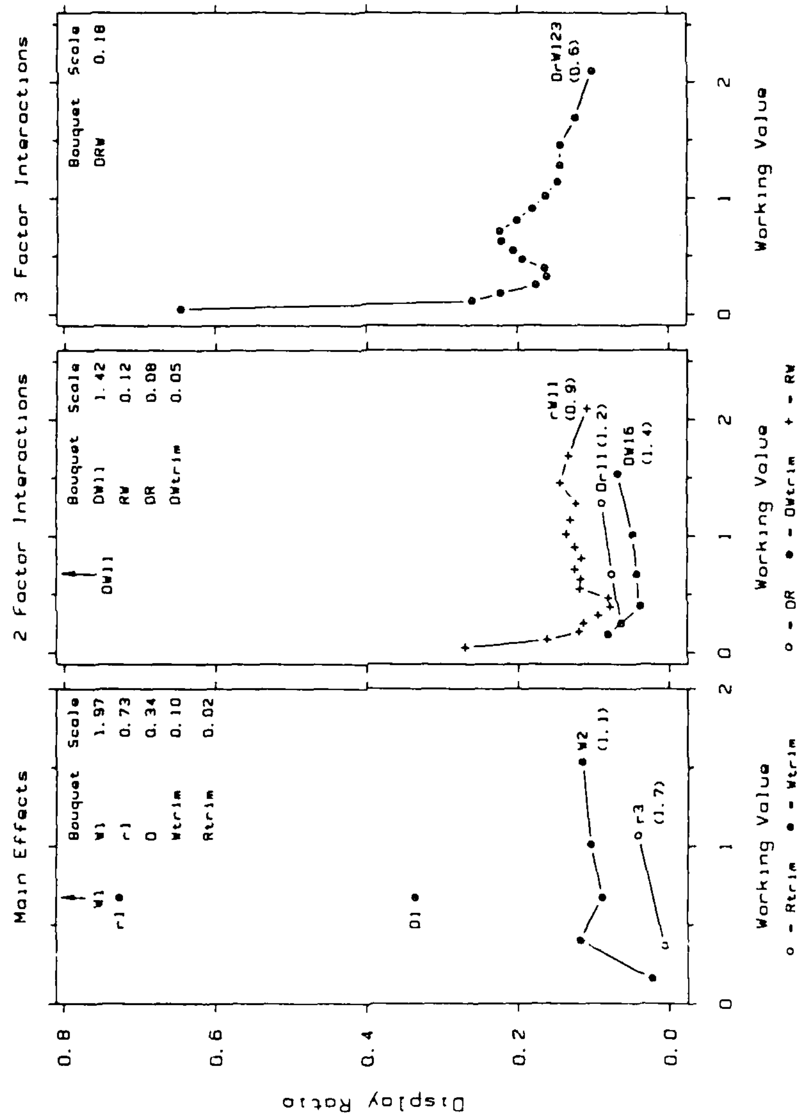


Figure 11. Person IB1 — log response time in Ln seconds. Helmert contrasts for rate — polynomial contrasts for weight. W1,DW11 nominated — r1 elected.

description of the data and has isolated the (main) effect of rate into the comparison of the response at the lowest rate with the average of the responses at all other rates.

D5. Refactoring

Our general attitude of "redoing anything that seems to deserve it, at least on a trial basis" should by now be clear. We have considered, in increasing order of drasticness: rescission into contrasts, re-expression of our response, (by implication at least) re-expression of our factors, and reformulation of the response. Well along in this order we should also consider another redo, refactoring of the pattern of analysis. Actually, as we shall soon see, our example already illustrates this.

Splitting

The earliest approach to the simplest sort of refactoring seems to be that of Brownlee's (1947) World War II concise book *Industrial Experimentation*, which was heavily concerned with 2^k designs with $k = 4$, 5 or 6 factors. Brownlee rightly, we feel, emphasized the frequent advantages of "splitting the experiment" and then analyzing and discussing the halves separately. This is particularly likely to help when the two versions of a factor were "do Q" and "don't do Q," and somewhat less likely to help when the versions were "the high level of Y" and "the low level of Y."

Brownlee decided whether or not to split in terms of the appearance of a significant interaction, which was not, for this purpose, compared with the mean square above it. We would feel that the proper reason for splitting is having something just above, the size of whose effect (or mean square) is not much larger than the interaction (which does itself need to appear not to be pure error). This formulation makes it much clearer what is to be split — those factors in the substantial interaction which do not appear in the label of the similar-sized mean square above.

In Brownlee's case — 2^k for small k, high-order interactions for error — it was not easy for an interaction to be significant, and if significance was reached, it was not easy for the main effects to be considerably still larger (unless they were known about all the time). Thus, for 2^k, his approach usually led to decisions to split that would also be made on the basis of what we assert to be appropriate reasons.

The fact that this would not be true for designs whose factors have several versions may account for the disappearance of the idea of splitting, both from Brownlee's later books and from the literature generally.

Splitting into Persons

The experiment from which our example is drawn was designed as a crossing of the sort of 2 × 4 × 7 treatment pattern we have been analyzing for a single person by a pattern involving 8 subjects. The subject pattern involved 2 persons in each cell of a 2 × 2 for male vs. female and seeing vs. blind. (The original failure to make a sensible analysis corresponded to an *a priori* assumption that replication of persons within cell belonged in the lowest error term, quite contrary to the trustworthy maxim that "people will be different!")

Actually, the 8 people did behave quite differently, both in slopes against individual factors, and in difference of slopes from day to day. Splitting, at least initially, the data into 8 portions, one for each person, seems to be an essential step in understanding what is going on. This is a simple and important instance of refactoring.

Once we have done this, we can look at sets of 8 numbers, one for each person, for both individual and collective responses, and ask what they seem to show, particularly in terms of the imposed 2 × 2 design. In general, we see little associated with the factors of sex and sight (somewhat confounded, as they were, with age) but strong emphasis on "people will be different".

We turn briefly to the question of seeking limited consistency of behavior across persons in Part E.

Tacit Refactoring

Actually, of course, the original data of Johnson and Tsao is best thought of as having already been refactored, perhaps unwisely, before anyone else ever saw any of it. The original 8 × 56 table, eventually given in Johnson's (1949) book, for 8 persons and 56 condition-date combinations, was a table of means, each of 5 individual trials. The order of trial of the 5 repetitions of 28 conditions for each person was stated to be randomized, though no details were reported. It would be strange indeed, in view of all the other things that appear to have been going on in this data, if there had been no time trends associated with the 140 = 5 × 28 trials for each of the 16 = 2 × 8 date-person combinations. The effects of these trends were buried, in an unknown but reputedly random way, in the table of means, which is all that later analysts have had at their disposal. In a real sense, this is also an instance of refactoring — if not of something still more drastic.

We believe, then, that this example illustrates — in more than one way — the need for, and importance of, refactoring as a standard part of an analyst's tool kit.

Splitting on "Date" in the Example

In the IB1 data, both the weight slope and the date-by-weight slope, if not nominated, would be elected. Indeed, the latter ($DW11$) is nearly as large as the former ($W1$).

The great difference in weight slopes for this subject from one day to the other is easily seen in a simple plot of means for date-weight combinations, as in Figure 12. This suggests splitting on date.

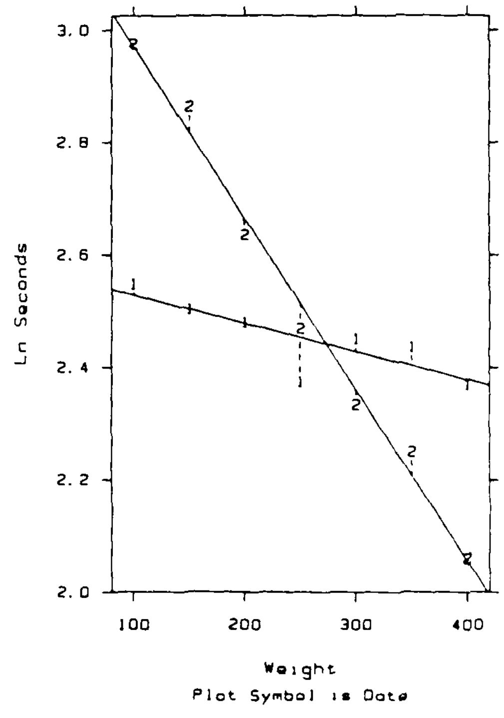


Figure 12. Person IB1 log response time vs weight separated by date (lines are the linear fit).

The classical analysis of variance, after nomination of all linear-to-the-j contrasts (now only for R, W, and RW) would take the form shown in Table 22 (using parallel columns for the two dates):

TABLE 22. Parallel-column analysis-of-variance table splitting on date.

	df		MS	
Label	(Date 1)	(Date 2)	(Rate 1)	(Rate 2)
Rate slope	1	1	.0322	.1130
Weight slope	1	1	.0704	2.6104
Interaction slope	1	1	.0737	.0022
Trimmed Rate	2	2	.0203	.0362
Trimmed Weight	5	5	.0066	.0059
Trimmed Interaction	17	17	.0114	.0200

It seems natural to present three panels of display ratios, one for rate, one for weight, and one for interaction, as in Figure 13, where we are treating each of the nominated contrasts ($d1R1$, $d1W1$ and $d1RW11$ for day 1, $d2R1$, $d2W1$ and $d2RW11$ for day 2) as separate one-contrast bouquets. The display ratio for the largest of these — $d2W1$, the weight slope for day 2 — is 2.4, which is 4.8 times larger than that of the next largest nominated contrast ($d2R1$).

Super-Election in the Split Example

Collecting the 6 nominated contrasts into a six-contrast nominated bouquet and then computing display ratios produces the first two columns of Table 23. Quite clearly, the weight slope for day 2 stands out from the rest. (The ratio-to-scale for $d2W1$ in the 6-contrast bouquet is 2.4.) Electing this contrast and post-trimming produce the last two columns of Table 23.

If, instead, we use the 3-contrast Helmert *SFP* bouquet as our scission of rate, we get the horizontalized plots of Figure 14. In this exhibit, we have nominated the linear weight-within-date contrasts ($d1W1$ and $d2W1$) and elected the two rate-50-versus-the-rest Helmert *SFP* contrasts of rate within date ($d1r1$ and $d2r1$). The election used a threshold value of 2. Notice that no rate-by-weight interaction can be elected with any reasonable threshold. As before, the use of the Helmert contrasts has concentrated the relationship between response and rate largely into a single contrast.

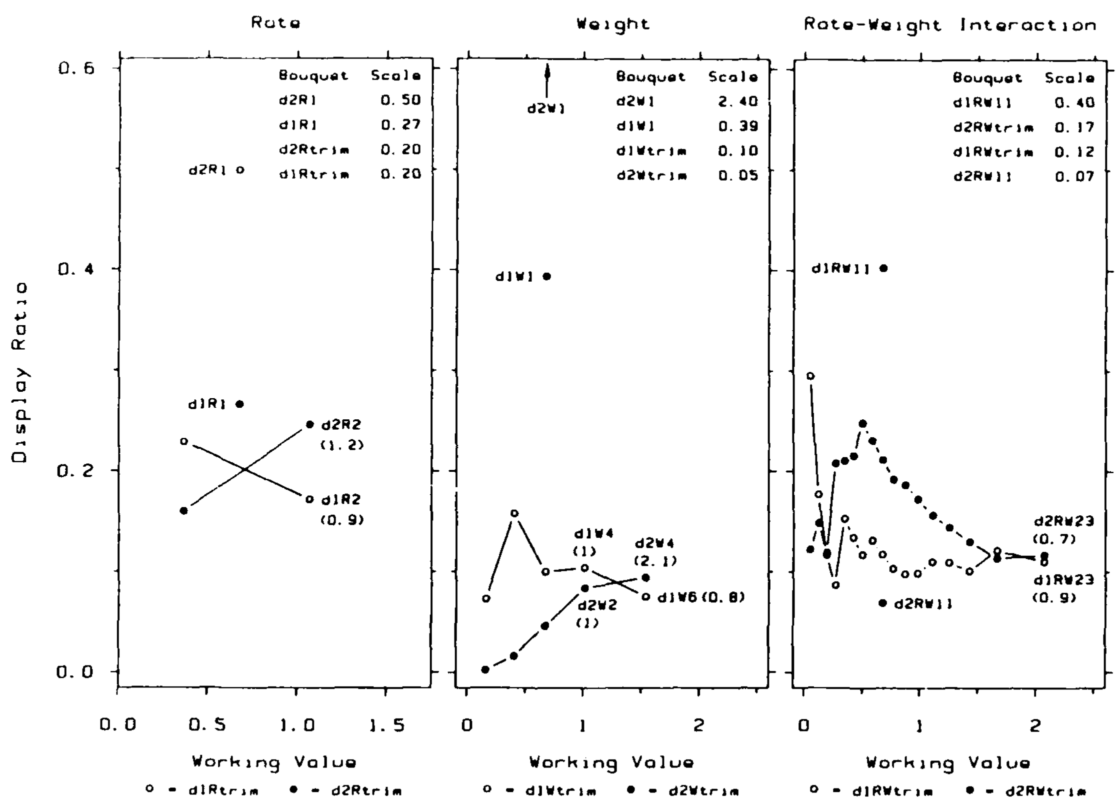


Figure 13. Person IB1 — log response time in Ln seconds. Splitting on date; polynomial contrasts for rate and weight within date.

Combining the two nominated and the two elected contrasts into a 4-contrast nominated-plus-elected bouquet, we get Table 24. The ratio-to-scale of $d2W1$, the linear weight-within-date-2 contrast, within the 4-contrast bouquet is 1.36, so that while this contrast is notable, it is not outstandingly large within the nominated-plus-elected bouquet. In the next section, we will give our final interpretation of the IB1 data, based on what we have now found.

TABLE 23. Display ratios for the nominated bouquet after splitting on date.

	6-Contrast Bouquet		Electing $d2W1$ and Post-Trimming	
Contrast	Display Ratio	Working Value	Display Ratio	Working Value
$d2W1$	.998	(1.620)	2.398	(.674)
$d2R1$	.300	(1.119)	.219	(1.534)
$d1RW11$	.337	(.804)	.268	(1.010)
$d1W1$	.477	(.555)	.393	(.674)
$d1R1$	.533	(.336)	.445	(.402)
$d2RW11$	.356	(.132)	.299	(.157)
	scale = .416		(trim) scale = .299	

TABLE 24. Display ratios for the nominated-plus-elected bouquet after splitting on date with Helmert contrasts for rate.

	4-Contrast Bouquet	
Contrast	Display Ratio	Working Value
$d2W1$	1.133	(1.426)
$d2r1$	.491	(.869)
$d1W1$	.530	(.502)
$d1r1$	1.366	(.194)
	scale = .831	

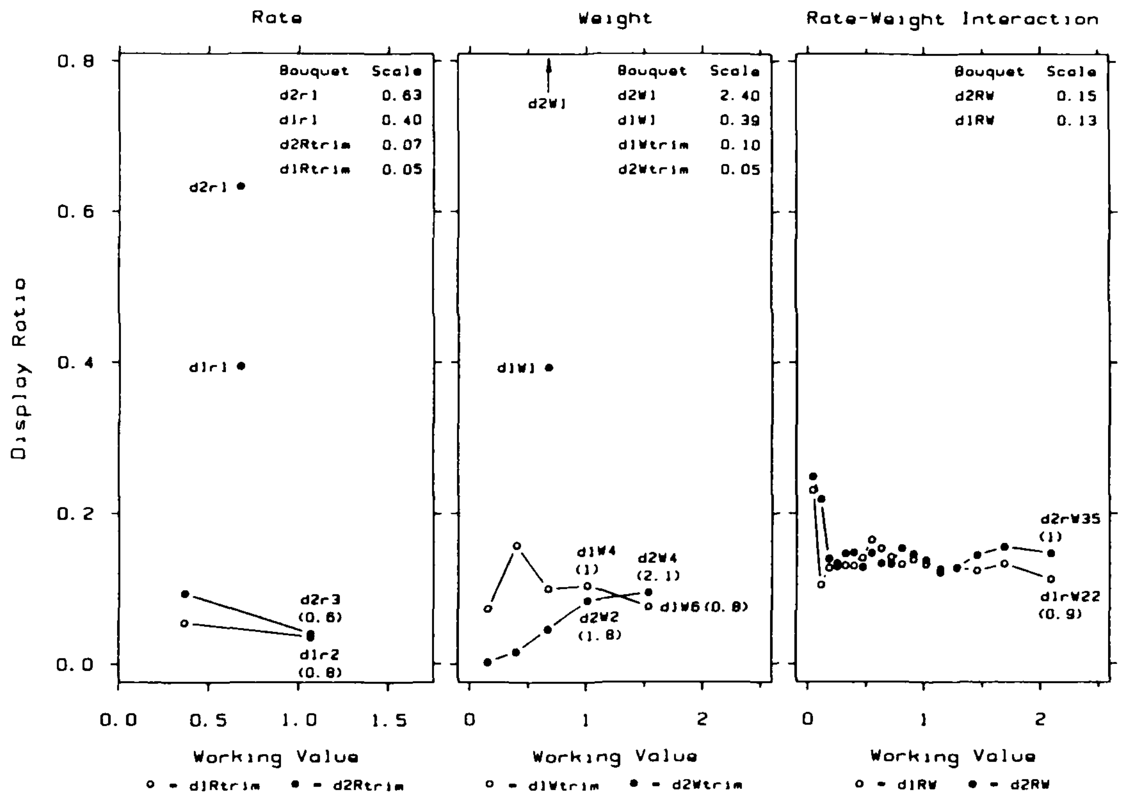


Figure 14. Person IB1 — log response time in Ln seconds. Splitting on date. Helmert contrasts for rate — polynomial contrasts for weight.

Two Further Illustrations

If we give no other example, the choice of the term "refactoring" may come into question. So we offer a few possibilities for 2×2 tables: one where an interaction becomes a factor and *vice versa*:

	Before	
	"no"	"yes"
to "no"	stay	move
to "yes"	move	stay

$\longrightarrow$ and

	Before	
	"no"	"yes"
stay	to "no"	to "yes"
move	to "yes"	to "no"

and one where an interaction is transferred into (a) modifying the A effect, (b) modifying the B effect, and (c) inserting a bonus in one chosen cell (can be any one of the four (cf. Seheult and Tukey, 1982)).

$$\left.\begin{array}{l}\text{A main effect}\\ \text{B main effect}\\ \text{AB interaction}\end{array}\right\} \longrightarrow \left\{\begin{array}{l}\text{A effect}\\ \text{B effect}\\ \text{A-and-B-both-high bonus}\end{array}\right.$$

D6. Recapitulation

Before considering the complete data set (of all 8 persons) and then summarizing the main thrust of this paper, it is likely to be helpful to recapitulate the steps we have taken, as seen in the light of where we now stand.

In Part A we discussed the analysis of the original $2 \times 4 \times 7 = 56$ numbers in terms that could, indeed should, have been planned in advance of seeing those numbers. Our analysis focussed on pictures corresponding to various analyses of variance, where the correspondence was mediated by (a) scission of each "line" into a bouquet of contrasts and (b) a horizontalized version, employing display ratios, of Daniel's half-normal plot. We did this for naive and linear-nominated analyses of variance, both conventional and aggregated. The linear-nominated analyses made clearer what the naive analyses suggested, namely that the essential description was in terms of a few linear-to-the-j slopes, a bending for low rates, and an apparently unstructured mass of residuals.

In Part B we considered election because of behavior in the data set before us, in contrast to nomination based on past experience. We found that much the same interpretations were suggested in our particular example by the results of election as by nomination, with the three largest of what would otherwise be the nominated contrasts, *R*1, *W*1, and *DW*11 being elected as important. We also considered the nominated bouquet consisting of all 7 linear-to-the-*j* contrasts and found that the rate slope *R*1 could reasonably be super-elected as the most important. By the end of Part B, we had come to the conclusion that, as far as the response of IB1 in grams was concerned, much of the relationship between the response and rate, weight and date could be adequately described by the six largest linear-to-the-*j* nominated contrasts: *R*1, *W*1, *DW*11, *DRW*111, *RW*11 and *D*1, in that order, where the linear rate slope is by far the most important.

In Part C we considered the possibility of reformulating the response and found that, since the response in grams was approximately proportional to rate, reformulating the response to log response time greatly reduced the apparent importance of the linear rate contrast and allowed other relationships within the data to become more apparent. In particular we discovered the relatively high levels of the display ratios for the trimmed rate bouquet and concluded that a systematic relationship between rate and the response was being spread across the polynomial contrasts.

In Part D we rethought the scission of rate into contrasts. After considering various bouquets which emphasized the ordering of the levels of rate, we concluded that the essential relationship between the response and rate was that person IB1 has a larger response for the smallest rate than the responses for the larger 3 rates, all of which are at the same level. By adopting the Helmert *SFP* scission of rate we found the apparent importance of all interactions involving rate vanished.

Lastly, we have considered the utility of refactoring of the pattern of analysis. We had already split the analysis by person based on the maxim that people are different. When we further split the analysis for person IB1 on date, we discovered that a strong slope in weight for day 2 predominated. Finally, we split over date and used the Helmert contrasts as our scission of rate. This eliminated the rate-by-weight interactions as potentially important.

At the end of our analysis of the response of person IB1 in log seconds, we are left with the following account of the data:

- a strong linear relationship between response and weight within day 2, with the response decreasing in weight. A much weaker, but still notable, linear relationship of the same direction between response and weight within day 1,

- a tendency for the responses at the lowest rate to be significantly higher than the responses at the other three rates, all three of which have responses at much the same level. This relationship is the strongest for day 2 but is still notable at day 1,
- a collection of very much smaller effects, interactions, and noise.

PART E: A LOOK AT THE OTHER 7 PEOPLE

E1. All 16 of the Person-Date Units

We have spent considerable effort on eccentric IB1. What of the other 7 people? With these choices:

response = log response time

no aggregation

nomination = all linear-to-the-j

a nominated bouquet

the horizontalized plots generally show relatively nice behavior. For 5 of the 7 people, the plots of the trimmed bouquets are relatively flat (except, sometimes, for the smallest contrasts) and all at roughly the same level. For these 5 people, the nominated contrasts contain the bulk of the information in the experiment, with little remaining in the trimmed bouquets besides background noise.

Two of the 7 people deserve specific attention: persons IB2 and IIA2. The horizontalized plots for these persons are shown as Figures 15 and 16.

Notice that in both of these figures the plots of the trimmed rate and trimmed date-by-rate bouquet appear at relatively high levels in relation to the levels of the bulk of the other trimmed bouquets. Furthermore, for each person and for each of the trimmed rate and trimmed date-by-rate bouquets, the slope of the line connecting the display ratios of the two contrasts within the bouquet is negative, indicating that the smallest contrast in the bouquet is relatively larger than might be expected. The form of these plots is symptomatic of the spreading of a systematic relationship across the contrasts in a bouquet.

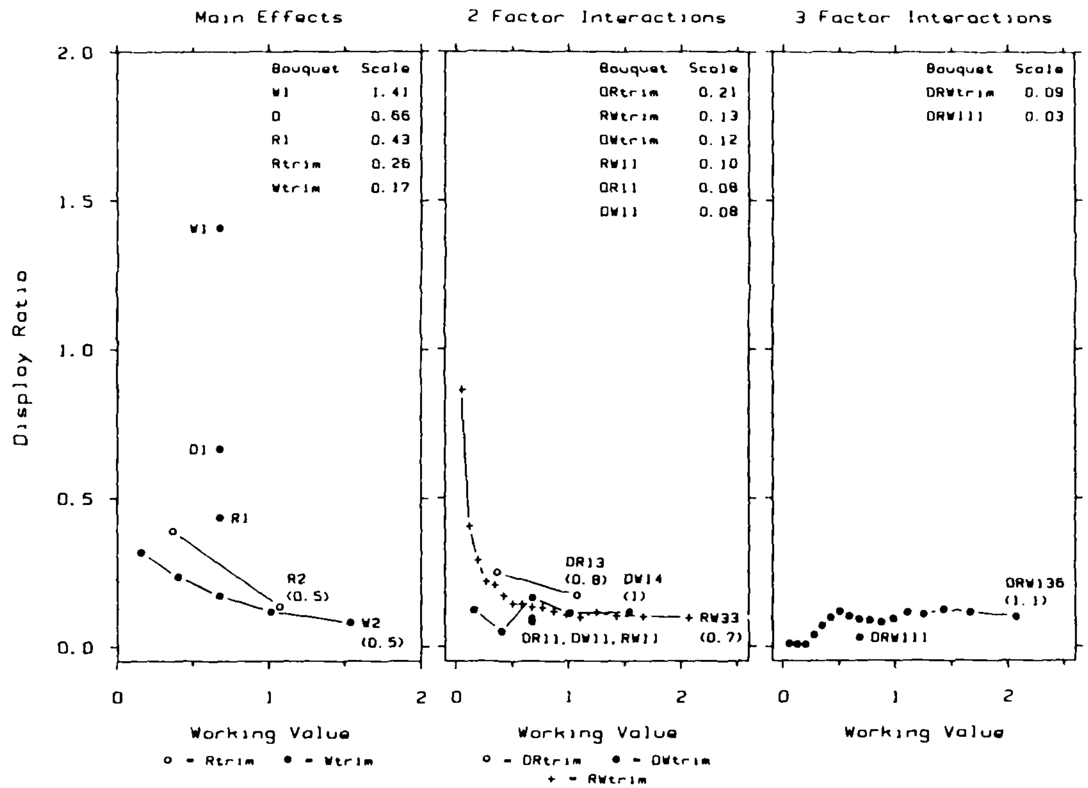


Figure 15. Person IB2 — log response time in Ln seconds. Polynomial contrasts — pretrimmed bouquets.

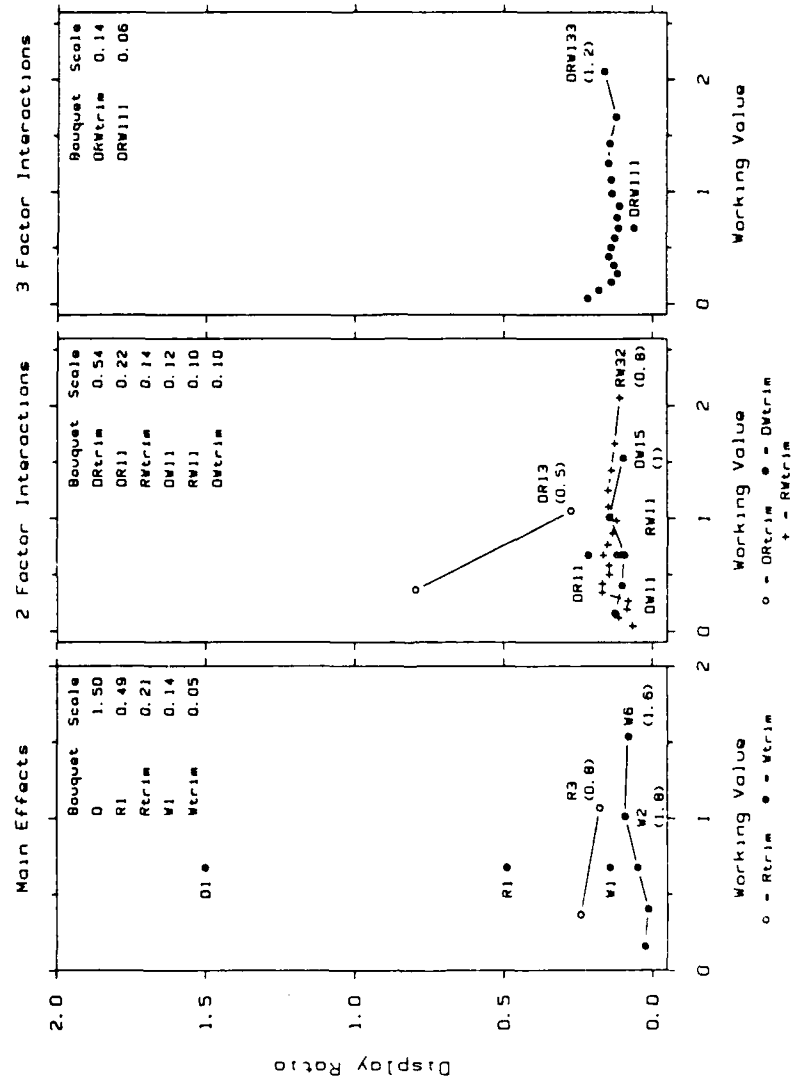


Figure 16. Person IIA2 — log response time in Ln Seconds. Polynomial contrasts — pretrimmed bouquets.

Figure 17 shows the type of spreading for each of the two people. In the plot of log response time versus rate within date for person IB2 shown as the first panel of Figure 17, we can see that the main systematic relationship that is not being well represented by individual polynomial contrasts is the large deviation of his response on date 2 to a rate of 150 gm/30 seconds from the line through the responses for the other three rates (for that date). This is a third-point-off deviation from linearity. In the second panel of Figure 17, there is a second-point-off deviation from linearity in the responses of person IIA2 on date 2. The reasons for these patterns of response for these two people, and the equivalent patterns for person IB1 (Figure 10), are unknown but demonstrate the maxim that people will be different. (We might be suspicious of the randomization.)

Figure 15 also shows a pattern in the display ratios for the trimmed weight bouquet. The monotonically decreasing nature of this plot is also indicative of spreading, of a less obvious kind, as indicated by Figure 18. The spreadings within the rate and weight bouquets are at least partly responsible for the high display ratios

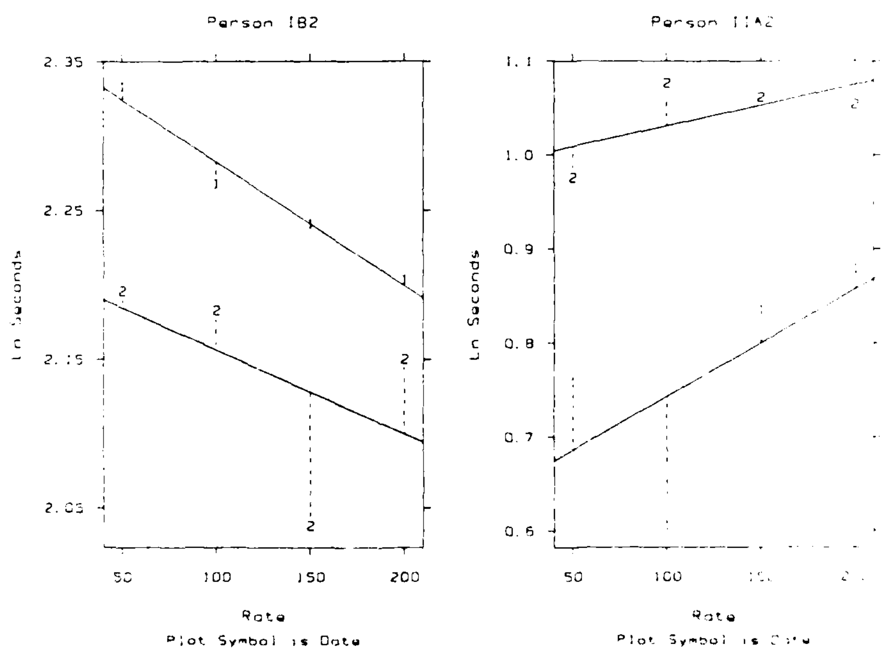


Figure 17. Log response time vs rate by date -- persons IB2 and IIA2.

seen in Figure 15 for the three smallest rate-by-weight contrasts (*RW*25, *RW*23 and *RW*16 in ascending order of size of contrast; descending order in display ratio).

Turning to the nominated bouquets for persons IB2 and IIA2, we note that the ordering of the contrasts is different between these two people and different from that of person IB1. The order of contrasts within the nominated bouquet for person IB2 is, from largest to smallest, *W*1, *D*1, *R*1, *RW*11, *DW*11, *DR*11 and *DRW*111, where *W*1 would assuredly be elected, *D*1 and *R*1 might also be, and the remaining 4 contrasts, all interactions, are at the background noise level. For person IIA2, the order of contrasts within the nominated bouquet is, from largest to smallest, *D*1, *R*1, *DR*11, *W*1, *DW*11, *RW*11, and *DRW*111, of which only *D*1 would be elected.

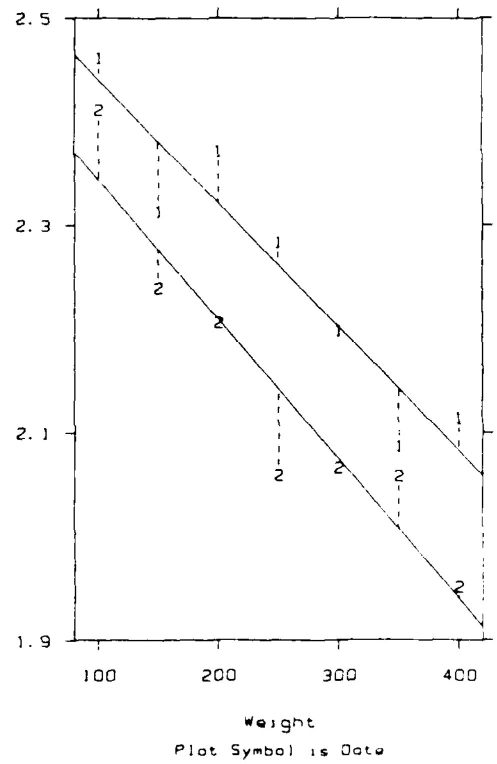


Figure 18. Person IB2: log response time vs weight separated by date.

Let Us Split Again

These analyses have split over person. If we further split over date and adopt a representation analogous to that used in section B5, making the various slope coefficients (and the centercept-multiplied-by-$\sqrt{28}$) equal to plus or minus the square root of the corresponding mean square, we get the results in Table 25. Also included in the table is the median of the 24 display ratios of the three trimmed bouquets within each date.

The major point to notice in Table 25 is that the relationships between response and the linear-to-the-j contrasts, split on date, tend to differ from person to person, and this is true even when we compare the two people of a specified sex and sight combination.

Any thoughts we have about the further analysis of Table 25 will have to wait for another day. If we go to working values of the square root of chi-square as divisors for the 16 observed median display ratios, we find that the picture looks best near 5 to 6 df in the chi-square. Clearly there is more variability here than we would expect for medians of 24 display ratios. We leave this, too, to another time, noting only the large apparent effect of sex × sight.

E2. Reassembly

Let us suppose, for simplicity, that for each of the 8 persons, or, perhaps, for each of the 16 person-date combinations, we have an analysis whose main constituents are:

1) a few fitted constants, say of the linear-to-the-j form for $j = 1, 2$ or 3 (for persons) or $j = 1, 2$ (for person-date combinations);

2) an unresolved mish-mash of residuals.

We understand how to look rather effectively at the sets of 8 or 16 individual responses under (1), or even, perhaps, at the 8 or 16 collective spreads extractable from (2), but how should we look at the 8 or 16 mish-mashes?

Better Matching

If we knew enough about the order of presentation of the treatments, which was once known, and if, also, the effect of order of presentation was somewhat consistent, across persons or across person-date combinations, we could match order of presentation across subjects (or across subject-date combinations) and see what could be extracted. It is probably reasonable to presume that such matching would be more effective than the only matching we can still try, namely that in terms of rate-weight-date or rate-weight combinations.

In either case, we can think of a two-way table of residuals, 8-by-56 or 16-by-28, where we seek to find some matching across persons, or across person-dates, but we do not want to — or dare not — assume that this matching runs right across all 8 persons or all 16 (person-date) combinations. These tables ought, it would seem, to be rescaled (for person or combination) before we tackle them. We need some form of *sub-factorial* analysis, since a plain factorial will not always work.

A first thing to do in such a situation would be to do a resistant row-PLUS-column analysis, say by median polish. If only a few persons, or a few combinations, escape some consistent pattern, such an analysis would tend to find the pattern.

If this fails, it seems natural to begin by dividing the 8 persons or 16 combinations into two or three clusters, and then repeating the first step for the smaller tables that result.

Delineations against Single Splits

Another, quite different approach, would be to select one split and plot the composite of the other splits against it, delineating (Tukey, 1977) the scatter. If the whole Johnson and Tsao data were split into 8 portions, one per subject, with a $2 \times 4 \times 7$ table of working residuals for each, this would mean plotting $7 \times 56 = 392$ residuals for other subjects against the 56 values of the selected subject. Enhanced by delineation, such a plot has a real hope of detecting commonality between the selected subject and a few of the other subjects. Eight such plots would not be too many to look at.

Bear in mind that the purpose of the present section is only to indicate that sub-factorial analyses are possible, and to stress that we need a body of experience in their use and modification.

TABLE 25. The results of splitting into 16 pieces (log response time coefficients are the signed square roots of the corresponding MS).

Label	Date	Centercept	R1	W1	RW11	Median Trimmed Display Ratio	Sight	Sex	Age
IA1	1	5.78	0.251	−0.023	−0.041	0.127	sighted	male	21,25*
	2	4.10	0.167	0.189	−0.093	0.060			
IA2	1	11.33	−0.200	0.423	0.116	0.097	sighted	male	25,21*
	2	10.89	0.014	0.114	−0.026	0.076			
IB1	1	12.98	−0.179	−0.265	0.271	0.114	blind	male	21,25*
	2	13.30	−0.336	−1.616	0.047	0.153			
IB2	1	11.97	−0.246	−0.630	0.032	0.105	blind	male	25,21*
	2	11.33	−0.167	−0.711	0.059	0.130			
IIA1	1	9.33	−0.382	−0.413	0.346	0.130	sighted	female	33
	2	8.90	−0.055	−0.285	0.053	0.118			
IIA2	1	4.08	0.338	0.126	−0.020	0.160	sighted	female	33
	2	5.52	0.131	0.011	−0.079	0.121			
IIB1	1	7.29	−0.062	−0.839	−0.022	0.092	blind	female	33
	2	6.33	−0.031	−0.528	−0.066	0.111			
IIB2	1	7.55	−0.049	−0.200	0.064	0.085	blind	female	33
	2	7.13	−0.050	−0.340	0.015	0.052			

* One blind male and one sighted male were 21, the other two males were 25.

PART F: SUMMARY

The overall thrusts of this account are:

1) That simple graphical views of what can easily be calculated from factorial data can guide us in structuring a useful description, one that:

    1a) gives detail, where that is useful,

    1b) avoids detail, where detail would clog our perception,

    1c) alters such parts of the initially concerned expression, formulation, and factoring as need to be changed for increased simplicity of description.

2) That scission of "lines" of an analysis of variance into bouquets of contrasts is a useful tool in doing this, noting that:

    2a) looking at alternative scissions can help, either by making visible something that deserves attention, or by increasing our confidence that we were not missing anything visible,

    2b) the choice of the basic scissions should be responsive to the nature of the corresponding factor: measured, ordered, weakly structured, or unstructured,

    2c) the use of product scissions for interaction lines is the best way we now know in which to begin (splitting or other refactoring may be urged on us by the results of the initial analysis),

    2d) for measured factors, the use of the conventional linear contrast as one of each of our bouquets seems desirable,

    2e) the most promising default may be to combine this contrast with the EPO contrasts (although combination with Helmert contrasts worked best for the rate factor in our example).

3) That modifications and enhancements of Daniel's classical half-normal plot, combined with (2), provide an effective way to do (1), noting that we can make our pictures easier to interpret and understand by:

    3a) "horizontalizing" the plot by plotting "display ratio" against working value, 3b) making separate, sometimes superimposed plots for different lines in the corresponding analysis of variance,

3c) separating lines, for the purpose of (3b), into as meaningful groups as we can find.

4) That enhancing the basic analysis into lines, by nominating *in advance* — or electing *because of the data's behavior* — certain contrasts to become separate "lines" in an enhanced or revised analysis can be very important, noting that:

4a) nomination is "safer" than election, and deserves our careful attention,

4b) the loss from uncalled-for nomination is small, since our procedures will often lead to reabsorption,

4c) the gain from needed nomination can be great, *both* in focussing more attention on the nominated contrast, when such focussing is needed, *and* in avoiding inappropriate dragging upward of the display ratios for other contrasts.

5) That a free hand in rescission, re-expression, reformulation, and refactoring can be of great assistance in such analysis, even though undeserved use of such freedom may produce, with increased frequency, simple descriptions of one data set that later prove not to extend to others, since:

5a) to trust in the establishment of either qualitative (e.g. structures) or quantitative (e.g. slopes) behavior by the analysis of *one* data set collected under *one* set of conditions is very poor science,

5b) the gain in understanding which usually accompanies a simpler description is so frequently helpful in the later use of the results, even when it reflects accidental serendipity.

We believe all of these points apply to the analysis of most factorial data sets with 3 or more factors (for the case of several factors at two levels see Seheult and Tukey 1982), and, as well, to a majority of other instances of analysis of variance of a similar size and complexity.

REFERENCES

Abelson, R. P. and Tukey, J. W. (1963). "Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order," *Annals of Mathematical Statistics*, **34**, 1347-1369.

Brownlee, K. A. (1947). *Industrial Experimentation*. New York: Chemical Publishing Company.

Chambers, J. M. (1977). *Computational Methods for Data Analysis*. New York: Wiley.

Daniel, C. (1959). "Use of half-normal plots in interpreting factorial two-level experiments," *Technometrics*, **1**, 311-341.

Daniel, C. (1985). "Patterns in residuals from a line through five points." Presented to the Northern New Jersey Chapter of the American Statistical Association May 6, 1985.

Green, B. F. and Tukey, J. W. (1960). "Complex analyses of variance: general problems," *Psychometrika*, **25**, 127-152.

Johnson, P. O. (1949). *Statistical Methods in Research*. New York: Prentice Hall.

Johnson, P. O. and Tsao, F. (1944). "Factorial design in the determination of differential limen values," *Psychometrika*, **9**, 107-144.

Kinderman, A. J. and Monahan, J. F. (1976). "Generating random variables from the ratio of two uniform variates." In *Proceedings of the Ninth Interface Symposium on Computer Science and Statistics*. Boston: Prindle, Weber & Schmidt, pp 197-200.

Pearce, S. C. (1953, revised 1976). *Field experimentation with fruit trees and other perennial plants*. Technical Communication No. 23 of the Commonwealth Bureau of Horticulture and Plantation Crops. Farnham Royal: Commonwealth Agricultural Bureaux.

Seheult, A. and Tukey, J. W. (1982). "Some resistant procedures for analyzing 2^n factorial experiments," *Utilitas Mathematica*, **21B**, 57-98.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

ATE

LMED

8